Non-native vowel processing as reflected by the brain: The Mismatch Response, the Acoustic Change Complex and Dynamic Causal Modeling

Georgina Louise Oliver

2013

Department for Speech, Hearing and Phonetic Sciences

A thesis submitted for the degree of Doctor of Philosophy

University College London

Primary supervisor: Prof. Dr. Paul Iverson

Secondary supervisor: Dr. Alexander Leff

Declaration

I, Georgina Louise Oliver, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

The aim of this thesis was to examine how auditory vowel processing by native and non-native speakers is reflected neuronally, specifically by the Mismatch Response (MMNm), the Acoustic Change Complex (ACC) and Dynamic Causal Modeling (DCM). We investigated whether these different brain responses represent different levels of speech sound processing and how the mismatch response is represented in the neuronal source architecture of native language (L1) and second language (L2) speakers.

In study 1, MEG MMNm data on L1 and L2 vowels was collected from English controls and French L1 speakers with a varying range of L2 proficiency. Additionally, subjects performed a range of behavioural tasks which targeted vowel perception (category discrimination and vowel identification) and production. The MEG data from this study was analysed conventionally and with dynamic causal modeling in order to determine neuronal sources and the dynamic source architecture in the L1 and L2 brain. In study 2, English controls and German subjects performed behavioural tasks (auditory discrimination and a combined assimilation/goodness of fit task). In study 3, EEG ACC data on L1 and L2 vowels was collected from L1 and German L2 speakers of English with a varying range of L2 proficiency. In study 4, EEG

ACC data on L1 and L2 vowels was collected from L1 and German L2 speakers of English with a varying range of L2 proficiency.

In summary, the MMNm indicated whether a speech sound had gained phoneme status in an L2. DCM showed that there is no difference architecturally and functionally between an L1 and a highly proficient L2 speaker's brain with regards to vowel processing. The right hemisphere supports the left during L2 vowel processing in low ability L2 speakers. The ACC was linked to individual vowel identification abilities, supporting the results from the DCM data.

Acknowledgements

Many people helped me on my journey towards the PhD, this is where I would like to thank some of you:

Paul Iverson, thank you for being the best supervisor I could have asked for. Alex Leff, thank you for being so patient and so ready to help. Tom Schofield, thank you for teaching me how to use the MEG machine, being my scan partner and explaining DCM to me whenever I asked. Sundeep Teki, thank you for helping me with the DCM analysis and supplying me with scripts. Emma Brint and Tim Schoof, thank you for teaching me how to use the EEG machine. Thank you to all my PhD colleagues for being so friendly and helpful – Yasna, Chierh, Nada, Wafa, Yasu, Dong-Jin, Katharine, Mark, Joaquin, Lucy, Hao, Kurt, Dorothea, Louise, Yue, Mauricio, Albert, Christiane, Gisela, Csaba, Young-Shin, Kayoko, Kota and Sonia!

My parents and my sisters Philippa and Caroline, thank you for everything.

My wonderful husband Holger Roth, your emotional support and Matlab skills were invaluable. I could not have done this without you.

INDEX

INDEX OF FIGURES

Chapter 1: Introduction

Non-native speakers tend to have a foreign accent when speaking a second language (L2). This is known to be due to native-language phonetic categories influencing second-language phoneme learning (e.g. Sebastien-Galles & Baus, 2005) as is evident from foreign-language accents. A well-known example for production difficulty in an L2 (Bradlow, Pisoni, Akabane-Yamada, & Tohkura, 1997; Hattori & Iverson, 2009; Strange & Dittmann, 1984) is the case of Japanese learners of English who have problems producing the English phonemes /r/ and /l/. These problems in using several phoneme systems could be because one phonemic category can contain several acoustically different sounds in one language while each of these sounds can belong to a different phonemic category in another language.

The difficulties in acquiring a new phoneme system can lead to L2 production difficulties – the aforementioned foreign accent. However, different phoneme inventories can also affect perception in an L2. Many speakers have problems discriminating between similar-sounding sounds in their non-native language (Cutler, Weber, Smits, & Cooper, 2004; Hattori & Iverson, 2009; Iverson & Evans, 2007, 2009; Iverson, Pinet, & Evans, 2011; Oliver, Gullberg, Hellwig, Mitterer, & Indefrey, 2012; Ringbom, 1992). L1 tuning affects the L2 in that individuals become more and more attuned

to their L1 phoneme system during very early infancy (Mechelli, Price, Noppeney, & Friston, 2003); six-month old children have already developed a phonetic specialization for their native language even before acquiring lexical items in their L1 (Kuhl, Williams, Lacerda, & Lindblom, 1992). These L1 tuning processes interfere with L2 learning (Hattori & Iverson, 2009; Iverson et al., 2011; Ringbom, 1992). It is however unclear at which speech processing levels these processes occur.

In my past research at the Max-Planck-Institute for Psycholingistics (Nijmegen, the Netherlands) I ran different cross-linguistic studies on L2 auditory sentence processing in noise. While using online measures such as reaction times (Oliver et al., 2012) and offline pen and paper measures such as cloze tests and other proficiency tests to investigate L2 auditory sentence processing, I was concerned with the fact that, even though reaction times are fast and online, they are still controlled to an extent by the participant and thus, ultimately, subjective. Consciously or subconsciously, the reactions (button-presses) to stimuli are mediated by, firstly, a reaction to an acoustic stimulus and, secondly, a conscious decision process of whether to press the button or not. While, in my opinion, still undoubtedly superior to offline pen and paper language testing, I nevertheless thought that there must be a more objective testing method than reaction times.

This method would not be influenced by decision processes of the participant and would offer a way to objectively and reliably test automatic language processing, not language processing which has gone through a behavioural filter first. Additionally, this method would be able to test how automatic language processing is in an L2 speaker. Offline pen and paper tests and online reaction time experiments do not test language processing automaticity. It is thought that the L2 language processing of low-proficiency L2 speakers is highly controlled and effortful and that high-proficiency L2 speakers have gone through a development from controlled towards more automatic processing (DeKeyser, 1997, 2001). Automatised processing is thought to involve improved processing speed and accuracy, unintentional and effortless processing, and, overall, more efficient online performance (DeKeyser, 1997, 2001; Segalowitz & Hulstijn, 2005; Segalowitz & Segalowitz, 1993; Segalowitz, Segalowitz, & Wood, 1998). Automaticity is a sign of proficiency in language processing and involves ballistic processing (Segalowitz & Hulstijn, 2005), that is, processing which is unstoppable once started. This improvement in automaticity is assumed to be caused by exposure to and use of the L2 and is thought to lead to a restructuring of formerly controlled global processes into automatised sub-processes which do not strain working memory and which operate in a ballistic manner (Favreau &

Segalowitz, 1983). So, in this sense, automaticity is related to skill improvement through practice and exposure (Newell & Rosenbloom, 1981) and leads to less effort in processing. In essence, I was interested in finding a testing method which, while automatic/obligatory and not under conscious control of the subject being tested, also tested how automatic L2 auditory speech processing was in a subject and at which level of automatic processing cross-linguistic differences between L1 and L2 speakers can be observed.

At this time I first heard of an automatic brain response, called the Mismatch Negativity/MMN (Näätänen, 1975), which is a brain response to perceived differences in a sensory environment and which is automatic and not generally thought to be mediated by attention, although there is a debate on whether the MMN is modulated by attention or not (Sussman, 2007). The MMN is seen to consist of two processes, one process in which patterns in a sensory enviroment are registered and a secondary process in which deviations from the established patterns in a sensory environment are detected. Experiments using the MMN which varied levels of attention seem to show that, while attention may possibly influence the establishment of patterns, it does not seem to influence the subsequent process by which deviations from established patterns are detected (Rinne, Antila, & Winkler, 2001;

Ritter, Sussman, Deacon, Cowan, & Vaughan, 1999; Sussman, Sheridan, Kreuzer, & Winkler, 2003).

I immediately thought that the MMN would be a good method to investigate automatic auditory L2 speech processing and gain more objective results which were not influenced by participant bias, motivation or, probably, attention. The reason I chose to use the MMN was that I was interested in using automatic auditory processing methods which are most likely to not be influenced by attention. I additionally chose the MMN paradigm as my first experimental method due to its extensive usage in the clinical field. As an experimental paradigm it has been proven to work consistently, even though all the mechanisms as to why it does work are not yet known.

The same search for a method to measure L2 auditory language processing, using objective and automatic methods, also led me to subsequently use another automatic brain response, the Acoustic Change Complex/ACC (Martin & Boothroyd, 1999), as another technique for investigating differences between L1 and L2 automatic speech sound processing. The ACC has an advantage over the MMN in that more response data can be collected in the same amount of time, making results more robust and allowing us to test more stimuli.

I was interested in using several different methods to tackle

the issue of L2 automatic speech sound processing. Both methods chosen have proven themselves to be robust and reliable in their respective clinical fields, however I was interested in finding out what these methods could tell us about an, as yet, relatively new field, namely L2 automatic vowel processing at the cortical level.

There are other methods which could have been used for investigating auditory L2 vowel processing, for example the Auditory Brainstem Response/ABR (Jewett & Williston, 1971). The ABR is a response at the subcortical level which originates in the brainstem and can be measured using electroencephalography from the scalp. Temporal and spectral characteristics of sounds are preserved in the ABR and ABR of complex sounds are mainly used to assess the functioning of auditory processing. Subcortical function has also been shown to interact dynamically with higher-level auditory processing (Skoe & Kraus, 2010).

However, research suggests that the ABR may be modulated by selective attention (Lehmann & Schönwiesner, 2014). Additionally, it may also not be particularly well suited to studies investigating differences between vowels, as some speech formants, including the second formants of many vowels, are above the range of brainstem phaselocking and may therefore not be visible in the response (Bidelman, Moreno, & Alain, 2013; Skoe & Kraus, 2010). As I was interested in using methods which are

attention-independent to look at L2 vowel processing, this method would not have been suitable.

Brain imaging techniques, while expensive and more complicated to set up and analyse than purely behavioural experiments, are an opportunity to explore L1/L2 tuning and automaticity issues at a different level and temporal resolution than traditional behavioural techniques. Brain imaging picks up automatic responses before the subject can give a standard behavioural response. It has emerged from previous studies (Dehaene-Lambertz, 1997; Dehaene-Lambertz, Dupoux, & Gout, 2000; Diaz, Baus, Escera, Costa, & Sebastien-Galles, 2008; Kaan, 2007) that brain responses offer a possibility to investigate low level auditory processing, not just at the level of auditory discrimination between non-speech sounds, but at a speech-specific level, as automatic brain responses are clearly affected by the language background of listeners. Automatic brain responses such as the Mismatch Response (commonly known as the Mismatch Negativity or MMN), which was first discovered by Risto Näätänen in 1975 (Näätänen, 1975), are understood to represent relatively low level perceptual processing which occurs automatically. There seems to be a difference between brain imaging responses and behavioural responses; cross-language magnetoencephalography (MEG) and electroencephalography (EEG) brain responses have been shown to

occur in advance of lexical processing in adults (Näätänen et al., 1997). This means that changes in brain structure and cortical organization can be explored before changes in behavioral measures are visible, opening up possibilities, for example, to explore L2 learning at an earlier point in time when learning might not yet be reflected in behavioural measures or to explore whether brain imaging measures are predictive of rate or outcome of L2 acquisition.

Other advantages of certain brain imaging techniques is that they not only can offer superior temporal resolution within the tens of milliseconds range (as with EEG or MEG), but can also be used to investigate the location of processes and the functional interconnectedness of brain regions during auditory processing (Friston, Harrison, & Penny, 2003).

With respect to L2 learning, current EEG evidence (Kaan, 2007) suggests that L2 speakers use the same processing mechanisms as L1 speakers, but these may be slower or used less or more depending on the degree of L1 proficiency or the aspect of language being investigated. However, there is also EEG evidence (Zevin, Datta, Maurer, Rosania, & McCandliss, 2010) which suggests that L1 speakers' brain responses to L1 consonants are left-lateralised while L2 speakers' brain responses to L2 consonants are right-lateralised.

The present studies explore at how low a processing level we can measure cross-language differences in vowel processing and what they relate to behaviourally. L1 experience is thought to influence only relatively late stages of neurological auditory processing (Dehaene-Lambertz et al., 2000; Näätänen et al., 1997). Obligatory auditory evoked potentials which were recorded from midline sites and elicited by syllables which varied with regards to voice onset time showed no effect of language experience (Elangovan & Stuart, 2011). In fact, these types of obligatory auditory evoked potentials are thought to only reflect acoustic feature processing of speech and are not thought to be influenced even by native language phonemes (Dehaene-Lambertz et al., 2000; Wagner, Shafer, Martin, & Steinschneider, 2013).

There are, however, studies which have investigated brain responses to auditory stimuli at an even earlier level, the subcortical level. These studies use the Auditory Brainstem Response or a part of it, the Frequency Following Response/FFR, and investigated pure tone stimuli (Jewett & Williston, 1971; Slabu, Escera, Grimm, & Costa-Faidella, 2010), pitch stimuli (G.M. Bidelman, J.T. Gandour, & A. Krishnan, 2011; Gavin M. Bidelman, J.T. Gandour, & A. Krishnan, 2011; Gockel, Carlyon, Mehta, & Plack, 2011; Lehmann & Schönwiesner, 2014) and categorical perception of vowels (Bidelman et al., 2013; Skoe & Kraus, 2010).

The Frequency Following Response/FFR is an evoked sub-cortical brain response which is elicited by continuous presentation of low-frequency tone stimuli. The FFR manifests as a periodic waveform which follows the individual cycles of a stimulus waveform.

Slabu et al. (Slabu et al., 2010) investigated the detection of infrequent novel pure tone stimuli by the brainstem. Infrequent changes in auditory frequency were detected as early as 30 ms post stimulus onset. They concluded that change detection of auditory frequency is a multistage process which starts in the brainstem and is then transmitted to higher cortical levels.

Using musical stimuli and pitch stimuli, Bidelman et al. (G.M. Bidelman et al., 2011) demonstrated that English L1 musicians and Mandarin speakers were more accurate at tracking pitch at the subcortical level than English L1 speakers. This study shows an effect of long-term experience with rapid pitch changes on the frequency following response in the brainstem, regardless of domain. However, it is questionable whether this effect is strictly the effect of speaking another language rather than the effect of extensive training in following rapid pitch changes – either as a musician or as a tone language speaker.

Several other studies investigating cross-language comparisons of the FFR show that speaking a tone language

natively enhances pitch encoding in the brainstem, irrespective of context. (Krishnan, Swaminathan, & Gandour, 2009; Krishnan, Xu, Gandour, & Cariani, 2005; Swaminathan, Krishnan, Gandour, & Xu, 2008)

In another study which further examined sub-cortical pitch processing with musicians, tone and non tone language speakers, Bidelman et al. (Gavin M. Bidelman et al., 2011) again investigated brainstem responses from Mandarin speakers and L1 English musicians and non-musicians. The participants performed a pitch discrimination task and listened to musical sequences while brainstem responses were recorded. While both the Mandarin speakers and the musicians showed stronger brainstem representation of the defining pitches of the musical sequences, behavioural and brain responses were linked only for musician participants, but not for the tone language speakers or non-musicians. They conclude that there is a dissociation between subcortical neural processing and the behavioural measures of pitch perception for the tone language speakers, indicating that sensory-level enhancement of musical pitch information only leads to cognitive-level benefits when it is of behavioural relevance.

In another study which investigated harmonic tones and pitch perception using the FFR, Gockel et al. (Gockel et al., 2011) concluded that the FFR reflects acoustic information present in the

stimulus, but does not directly represent pitch perception itself.

All these studies investigated tones and pitch effects and showed the differences between tone language and non tone language speakers in brain stem responses and the FFR. However, it would be interesting to know whether there are any other language effects apart from pitch on subcortical processing. The effect of pitch could possibly even be said to be language-independent in the sense that it reflects the long-term experience of behaviourally meaningful pitch processing, either in a linguistic or in a musical context. This possibility is supported by Hickok & Poeppel (Hickok & Poeppel, 2004), who argue that early auditory processing stages in the brainstem may process acoustic data in such a way that is relevant to both linguistic as well as non-linguistic auditory perception.

A more complex or higher order of processing would be that of speech sounds. In order to investigate speech sounds and the way they are processed in the brainstem, Bidelman et al. (Bidelman et al., 2013) conducted a vowel categorical perception study with simultaneous brainstem and continuous cortical ERP recording and behavioural measures. The brainstem ERP reflected the acoustic properties of the stimulus (i.e. formant change). However, brainstem and cortical ERP recordings lacked any correspondence, despite behavioural and cortical ERP showing evidence of

categorical perception. These results indicate that brainstem responses and cortical responses are dissociated, so that the brainstem responses are not, in fact, a direct correlate of vowel perception. In fact, Bidelman et al. consider it probable that the brainstem is too low-level cortically to be able to provide the abstract phonetic representations which are necessary for categorical perception. According to Bidelman et al. (Bidelman et al., 2013), correlations between cortical brain responses and categorical speech percepts are visible within the first few hundred milliseconds post stimulus onset and emerge no earlier than late primary or secondary auditory cortex. Comparable results have been obtained from epileptic patients during intracranial recordings which indicate that the neural correlates of categorical perception are generated within the superior temporal gyrus (Chang et al., 2010; Mesgarani & Chang, 2012).

According to Krishnan et al. (Krishnan & Gandour, 2009), there are experience-driven adaptive neural mechanisms involved subcortically in the processing of pitch contours. These adaptive mechanisms enhance the response properties of neurons which are tuned to process those pitch contours which are relevant to the prosodic needs of a specific language. As such, according to Krishnan et al., these neural mechanisms are processing particular acoustic features which are common to speech rather than speech

itself.

In sum, when looking at sub-cortical auditory processing, there seems to be a beneficial effect of long-term experience in processing pitch musically on performance in pitch discrimination. Additionally, tone language speakers are better at processing pitch than non tone language speakers. However, when looking at categorical perception of vowels, there is a dissociation between brainstem responses and behavioural discrimination ability which indicates that the brainstem is too low-level to process abstract phonetic representations.

Taken together, the evidence so far points to the Auditory Brainstem Response and Frequency Following Response not being speech-specific with respect to pitch, but experience-specific, which is why musicians and tone language speakers are better than non tone language speakers. Being a musician and tone language speaking both benefit from brainstem frequency-tracking abilities which are tuned over time through exposure.

In order to explore how speech sounds are processed in the brain at a higher, cortical, level, we investigated two different obligatory auditory evoked cortical potentials: The Mismatch Response and the Acoustic Change Complex. The mismatch response, commonly known as the MisMatch Negativity/MMN, is an obligatory neuronal response to an unexpected sound (called a

deviant) in a train of identical sounds (called standards). The mismatch response can be measured either by MEG (magnetoencephalography) or EEG (electroencephalography).

In Chapter 2 we investigated the mismatch response (using MEG) with L1 English and L1 French/L2 English speakers. Subjects additionally performed in behavioural perception and production tasks. In Chapter 3 we focused on the neuronal network architecture underlying the mismatch response for L1 and L2 speakers and looked at the links between neuronal architecture and behavioural tasks. We analysed the MEG mismatch data using Dynamic Causal Modeling/DCM. DCM is a method of modeling the neuronal network underlying the auditory processing of native and non-native language speech sounds. In Chapter 4 we investigated the very early obligatory auditory elicited neuronal response called the Acoustic Change Complex, the ACC, using EEG with L1 English and L1 German/L2 English speakers. The ACC is an obligatory response to acoustic changes within a continuous chain of rapid auditory stimuli. Additionally, for both the MEG and the EEG experiments, we collected behavioural data from the subjects and looked at the relationship between brain and behavioural measures.

We were interested in finding out whether the mismatch response and the acoustic change complex show an effect of native and non-native language, how these changes might be expressed in

auditory neuronal architecture and in elicited obligatory MEG and EEG auditory evoked potentials and how these brain responses relate to the behavioural measures collected.

Chapter 2: The Magnetic Mismatch Response

Introduction

One well-known brain imaging response which has been used in many medical and language studies for the past 35 years is the mismatch response (Näätänen, 1975, 2007; Näätänen, Gaillard, & Mäntysalo, 1978; Näätänen, Paavilainen, Tiitinen, Jiang, & Alho, 1993; Näätänen & Picton, 1987). The mismatch response is a component of event-related potentials (ERPs), measured by electroencephalography (called the mismatch negativity/MMN), or event-related fields (ERFs), measured by magnetoencephalography (called the magnetic mismatch negativity/MMNm or mismatch field/MMF). Electroencephalography (EEG) measures the electrical fields generated by the activity of very large populations of neurons in the brain. Magnetoencephalography (MEG) measures the magnetic fields which are generated by the electric current flow in large neuron populations of at least 50000 neurons.

The mismatch response is not directly visible in raw EEG/MEG data, but must be computed, after pre-processing the raw data, by subtracting average brain responses to experimental stimuli from average brain responses to non-experimental stimuli. This component is generally found at about 150-250ms post-stimulus and, with EEG, its peak is negative-going. The mismatch response

is elicited automatically and can be recorded without the attention of the subjects. In fact, most studies using the mismatch response instruct subjects to read a book (Winkler & Czigler, 1998), look at pictures (Leff et al., 2009; Schofield et al., 2009), watch a silent movie (Jacobsen, Schröger, & Alter, 2004; Kasai et al., 2001; Peltola et al., 2003; Pulvermüller et al., 2001) or perform a visual tracking task (Winkler, Karmos, & Näätänen, 1996b) while brain data acquisition is running.

A common experimental paradigm is the oddball paradigm (Leff et al., 2009; Näätänen, 2001, 2007; Näätänen, Paavilainen, et al., 1993; Näätänen, Schröger, Karakas, Tervaniemi, & Paavilainen, 1993; Peltola et al., 2003; Pulvermüller, 2003; Pulvermüller et al., 2001; Schofield et al., 2009; Winkler & Czigler, 1998), in which a sequence of sound stimuli is presented auditorily with deviant sounds interspersed intermittently. The brain responds to the unexpected deviant sounds (the "oddballs") by producing a mismatch response. The mismatch response is thought to indicate pre-attentional processing at an automatic processing level and to show the brain's automatic change-detection response to changes in an unattended auditory environment and its subsequent attentional switch to the deviant sound (Näätänen, Paavilainen, et al., 1993; Näätänen & Picton, 1987). This attentional switch is indicated by the P3a component often following the mismatch

response and other physiological signs correlated with attention such as heart rate deceleration and skin conductance changes (Winkler & Czigler, 1998).

According to Näätänen et al. (Näätänen, Paavilainen, et al., 1993; Näätänen, Schröger, et al., 1993), repeated exposure to stimuli results in the development of short-term neural representations or memory traces which are stored in the brain: The auditory properties of the deviant sound are compared with a memory representation of previously heard sounds and are found to be different from it. The mismatch response is not elicited by sounds presented in isolation (Cowan, Winkler, Teder, & Näätänen, 1993) as the response needs established standards for comparison to deviants. In fact, at least one successive repetition of a standard is necessary for successful mismatch response elicitation (Jääskeläinen et al., 2004; Näätänen, 2001). Amplitude and latency of the mismatch response varies depending on the degree of deviancy of the deviant (Peltola, Tuomainen, Koskinen, & Aaltonen, 2007): the more deviant the stimulus is perceived as being from the standard, the higher the amplitude and the shorter the latency.

There are differing views on the exact nature of the mismatch response. The traditional view (Näätänen, Paavilainen, et al., 1993) of the mismatch response sees it as an automatic and uncontrolled response to auditory changes in the environment which depends

upon short-term sensory memory representations and is followed by an attentional switch. This view has been challenged by new findings (Garrido, Kilner, Kiebel, & Friston, 2007; Garrido, Kilner, Stephan, & Friston, 2009; Jacobsen et al., 2004) which have emerged from experiments using speech stimuli, instead of simple or complex tone stimuli. Most mismatch response studies so far (Näätänen, Schröger, et al., 1993; Winkler, Cowan, Csepe, Czigler, & Näätänen, 1996; Winkler & Czigler, 1998; Winkler, Karmos, et al., 1996b) have used simple or complex tones in order to reach conclusions about auditory processing and native language representation in the brain. However, this use of simple or complex tones is now seen as probably being too simplistic for language-related research as tone stimuli are not representative of a realistic speech environment. It may also be possible that non-speech stimuli might not be represented in the brain in the same way as speech stimuli.

Therefore, mismatch response studies (Cheour et al., 1998; Dehaene-Lambertz, 1997; Dehaene-Lambertz et al., 2000; Diaz et al., 2008; Gao et al., 2012; Jaramillo et al., 2001; Kasai et al., 2001; Näätänen, 2001; Näätänen et al., 1997; Pulvermüller et al., 2001; Teki et al., 2013; Uther, Giannakopoulou, & Iverson, 2012; Winkler, Kujala, et al., 1999; Winkler, Lehtokoski, et al., 1999) have started to use speech sounds as stimuli. This is especially

useful for exploring the representation and processing of a non-native language in the brain, thus allowing not only the comparison between two different tones, differing only in acoustic properties such as pitch, duration or intensity, but also between sounds from two different languages. Speech per se can be seen in two possible ways: A kind of specialized auditory input which is distinct from other complex non-speech sounds or as a type of complex auditory stimulus which is processed in a similar way to other types of complex auditory stimuli (such as music).

There arises the question as to whether the brain recognises speech stimuli as somehow different and processes them more efficiently than non-speech sounds or whether all stimuli are seen as equal and merely differ in terms of physical features (such as duration, pitch or intensity). In order to explore this question, Jaramillo et al. (Jaramillo et al., 2001) used an oddball paradigm with vowels and complex tones to elicit an MMN. The findings indicated that speech sounds (e.g. vowels) are indeed processed differently than complex tones. Vowels in general were more sensitive to durational changes and elicited a larger amplitude MMN and larger P3a components (although the sensitivity in duration in this case could also have been due to the language of the experiment, which was Finnish, a language where durational changes in vowels are meaning-bearing).

Additional evidence supports this view: Using an oddball paradigm, between- and within-category vowels were compared to tones (Kasai et al., 2001). It was of interest whether the hemispheric laterality of the mismatch response was different for each of these types of sounds. Between-category vowels elicited a left-hemisphere predominant pattern while within-category vowels and tones elicited right-hemisphere predominant patterns, indicating not only that within-category and between-category vowel distinctions are processed in different ways in the brain, but that within-category vowel distinctions may not even be classed as speech-specific. Different types of physical changes and features are thus not processed in the same way in speech and non-speech sounds, indicating that language is in some way special and differently represented in the brain.

This specialness of language is also indicated by findings that non-speech stimuli evoke right-hemisphere dominant responses (Levanen, Ahonen, Hari, McEvoy, & Sams, 1996) while speech stimuli evoke left-hemisphere dominant responses (Näätänen et al., 1997) and the question of whether brain responses to auditory stimuli are dominant to the right or the left hemisphere or equal in both, ie. hemispheric differences in investigated in many studies. Studies using the mismatch paradigm and EEG or MEG have used non-speech stimuli such as pure tones (Levanen et al., 1996;

Teamu Rinne et al., 1999), spectral deviants (Okamoto & Kakigi, 2013), lexical tones (Luo et al., 2006) and musical chords (Tervaniemi & Hugdahl, 2003) and evoked right-hemisphere dominant responses with these. Speech stimuli using the same paradigm, on the other hand, tend to evoke evoke left-hemisphere dominant responses (Naatanen, Paavilainen, Rinne, & Alho, 2007; Näätänen et al., 1997), using consonants (Luo et al., 2006) vowels (Teamu Rinne et al., 1999; Tervaniemi & Hugdahl, 2003) and temporal deviants (Okamoto & Kakigi, 2013). This is thought to be due to the right hemisphere being specialised to process spectral differences, such as lexical tones and music, and the left hemisphere being specialised to process temporal differences, such as speech-sounds (Okamoto & Kakigi, 2013).

Additionally to results from conventional MMN/MMF analysis, the computed neuronal source of the mismatch response to speech sounds is dominant in the left hemisphere (Friston et al., 2003). The use of speech stimuli as opposed to non-speech stimuli has led to a new view of the mismatch response. There has been a shift from identifying the function of the MMN-generating process as primarily related to the acoustic detection of deviant stimuli to seeing the process as being linked to the representation of auditory context, namely the standard. According to this view of the mismatch response, the detection of deviant sounds may not in fact

be the primary function of the mismatch response generating process. Its main function seems to be the continuous adjustment of a neural model of the auditory background.

The mismatch response is known to be indifferent to attentional focus. This means that patterns of auditory input are detected automatically and organized in a dynamic neural model. The automatisation of detection enables the auditory system to manage the majority of the sensory input without stressing the limited resources of the attentional system. The auditory neural model requires constant maintenance (Winkler, Karmos, et al., 1996b), whenever previously detected patterns and regularities are violated then the related parts of the model have to be adjusted accordingly (Jaramillo et al., 2001).

This idea of a dynamic neural adjustment model makes the mismatch response more interesting for language research. Instead of being merely a brain response which signals some kind of additive or subtractive feature change in the auditory environment, the new hypothesis allows the mismatch response to be interactive and it is thus possible to imagine how an automatic neural representation of language at several linguistic levels might be built up through experience and exposure to language sounds and structures.

Stimulus type is not the only factor known to have an

influence upon the mismatch response. Individual differences between speakers (native or non-native speakers) also could have an effect on the mismatch response. Individual differences in the ability to process non-native speech sounds do seem to exist. These differences could be due to physical characteristics of the brain. An example for this is delivered by a study in which faster learning of non-native Hindi contrasts was linked to individual differences in brain structure in anatomical MRI (Golestani, Molko, Dehaene, LeBihan, & Pallier, 2007; Golestani, Paus, & Zatorre, 2002). This suggests that anatomical characteristics of an individual's brain, such as white matter volume and symmetry, are linked to faster learning of a non-native L2 phoneme. It is however not clear in which way learning and physical characteristics are linked with each other and whether there is a causative relationship between brain structure and phoneme learning or whether there is another variable which is driving this relationship. When looking at functional brain measures, not brain structure, a training study (Näätänen, Schröger, et al., 1993) explored the link between behavioural L2 and functional brain measures and found differences between individuals both initially and after completing discrimination training on non-native speech sounds.

The mismatch response may therefore possibly be useful in order to predict ultimate non-native language learning success,

non-native language learning aptitude and the effect of sound discrimination training on higher level linguistic knowledge acquisition. Previous evidence from a recent study (Diaz et al., 2008) has pointed towards higher proficiency/aptitude correlating with higher MMF and MMN amplitudes; this study (Diaz et al., 2008) showed a larger MMN amplitude for native language vowels (L1=Spanish) for individuals who were better at perceiving L2 vowels (L2=Catalan). The results of this study seem to indicate that a larger MMN amplitude might correlate with a higher proficiency in L1 and L2 auditory language processing.

However, one might ask what learning an L2 speech sound means. Is learning an L2 category merely one specific ability which can, if necessary, be trained and improved with a specific task? Or is it in fact more likely that learning an L2 speech category is based on a range of skills which need to be trained in parallel? Evidence for the latter theory is demonstrated by studies (Iverson & Evans, 2007, 2009; Iverson, Hazan, & Bannister, 2005; Iverson et al., 2011) which employ batteries of perception and production tasks and see whether these are linked to one another pre- and post-training. The perception of non-native speech sounds involves many complex underlying abilities which need to be developed in a second language. The different abilities needed to perceive and produce a second language are not necessarily related to one another and it is

not yet clear how these underlying abilities in perception and production correlate with each other and with automatic auditory brain responses such as the mismatch response. L2 speakers who are better at identifying a non-native speech sound are not necessarily better at producing it and vice versa (Hattori, 2009). There is a large amount of evidence from previous studies (Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; Bradlow et al., 1997; Fu, Galvin, Wang, & Nogaki, 2005; Hazan, Sennema, Iba, & Faulkner, 2005; Lambacher, Martens, Kakehi, Marasinghe, & Molholt, 2005; Logan, Lively, & Pisoni, 1991) which points towards a link between production and perception.

Auditory training seems to improve the efficiency of categorization (reflected in vowel identification/ID scores) in an abstract way, making it generalisable to novel stimuli, whilst not simultaneously improving the sensitivity (reflected in category discrimination/CD scores) to an L2 sound. This is seen in a recent study by Iverson et al. (Iverson et al., 2011) with French listeners who were being trained on English vowels improved their vowel identification (ID) abilities over time, but did not improve much in their category discrimination (CD) abilities.

Studies have found correlations between perception and production of L2 phonemes (Bradlow et al., 1997), between L2 vowel discrimination and L2 vowel production abilities (Flege, 1999)

and L2 consonant perception and production (Hattori, 2009). However, evidence from training studies (Hattori, 2009; Iverson et al., 2011) indicates that perception and production abilities in an L2 are not in fact particularly strongly correlated. Perceiving and producing an L2 phoneme seems to involve different abilities, as even subjects who had been trained to a high level in perceiving and producing L2 phonemes did not show a strong correlation between tasks involving L2 perception and production. Another study (Golestani & Zatorre, 2009) examined the relationships between identification, discrimination and production of Hindi consonants: While the identification and discrimination of non-native Hindi dental-retroflex contrasts were linked to each other, Hindi consonant production abilities varied widely between individuals.

The present study explored both behavioural measures and brain measures, specifically the mismatch response as measured by MEG, as an indicator for the ability to distinguish between L2 phonemes and ultimately the representation of an L2 phoneme in the brain. This experiment investigated how perceptive and productive L2 abilities are linked to each other by measuring individual accuracy in between- and within-category L2 phoneme distinctions, automatic brain responses and L2 production. French native speakers have difficulties differentiating between the English

phoneme /i/ (as in "beat") and /ɪ/ (as in "bit") as /ɪ/ is not a French phoneme (Iverson et al., 2011). For this reason this experiment looked at how French native speakers perceive, discriminate and produce these difficult English phonemes, as variability in perception, discrimination and production proficiency between individuals can be expected due to the choice of our subjects who varied in their L2 proficiency.

This study explores how L2 speakers discriminate, perceive and produce vowels belonging to either the same or different non-native categories and which of these measures are linked most strongly with the mismatch response. The ability to perceive and produce within-category and between-category distinctions is assumed to vary with differing L2 proficiency. The present study measures outcomes from an array of behavioural tasks which are assumed to tap into underlying auditory speech-sound processing abilities at three different levels of speech-sound perception: the perceptual level (vowel identification), the phonetic encoding level (category discrimination) and the auditory perception level (auditory discrimination). Additionally, subjects performed an L2 production task, were measured using MEG and completed a short language background questionnaire which included questions on age of acquisition and length of residence in the UK.

Method

Subjects

All subjects were tested in London (UK). Two groups of adults participated in all of the five tasks – English native controls (English L1) and non-native speakers of English (French L1, English L2 speakers). The native controls consisted of 9 adult native speakers of English (3 females, 6 males). Their age ranged from 23 to 43 years at test (mean= 28.11 years). All controls were brought up in a monolingual environment in England. All of the controls were right-handed and reported no hearing problems. The L2 speakers of English consisted of 13 adult native speakers of French (9 females, 4 males). All participants resided in London at the time of the experiments and were tested in London. Their age ranged from 22 to 40 years at test (mean= 27.4 years). Their age of acquisition of English as a second language through formal education at school ranged from 7 to 16 years (mean= 11.3 years) and length of residence in the UK ranged from 1 month to 9 years (mean= 32.3 months, SD=27.8). 3 subjects had lived in other English-speaking countries, length of residence in other English-speaking countries ranged from 24 to 43 months (mean=32.3 months). All subjects were brought up in a monolingual environment in France. All of the

subjects were right-handed and reported no hearing problems. All subjects (L1 and L2 speakers) participated in all experiments.

Stimuli

The stimuli for vowel identification and category discrimination consisted of natural recordings of the vowels /i/, /ɪ/ and /u/ in a b-V-t context spoken by ten speakers of British English, five female and five male. These stimuli were a subset of the test stimuli used in previous studies (Iverson & Evans, 2009; Iverson et al., 2011). The perceptual experiments were all conducted in a quiet room. A laptop was used to play the stimuli and collect responses from the subjects. Stimulus recordings were made in an anechoic chamber.

The stimuli for auditory discrimination and the MEG experiment both consisted of synthetic words. The auditory discrimination (AD) task consisted of synthesised words based on best exemplar stimuli from a previous study by Iverson et al. (Iverson et al., 2011). The stimuli were created using the cascade branch of a Klatt synthesiser (Klatt & Klatt, 1990). In this study, however, vowel duration was equated, the F3 frequency was not varied (F3=3200Hz), and there was no formant movement. The words consisted of variants of the vowel /i/ in /b/-V-/t/ contexts: /bit/ and one variant /bit/ item. The two synthetic vowels /i/

differed from each other in F1 and F2, were however equalised as to duration, RMS (root mean square) and F3. The variant differed by 1.5 equivalent rectangular bandwidth/ERB (Glasberg & Moore, 1990; Moore, Glasberg, & Bear, 1997) from the standard /i/. The stimuli were designed to be within the /i/ category. The variant had an increased F1 (F1=247Hz) and a decreased F2 (F2=2479Hz) compared to a standard /i/ realisation.

This particular ERB value was chosen on the basis of an earlier pilot study with several ERB values with 20 French L1 speakers (13m, 7f) aged 17 to 52 years (mean=25.6, SD=8.2). Age of acquisition of English as an L2 ranged from 8 to 16 years (mean=11.8 years, SD=1.8) and their length of residence in the UK ranged from 1 month to 23 years (mean=33.7 months, SD=62). The pilot consisted of an oddity paradigm with four vowels which were all deviants of the non-native phoneme category /i/. One vowel was a standard /i/, the other three vowels were synthesised within-category deviants of /i/ which differed as to ERB distance from the standard /i/. Deviant 1 differed by 0.75 ERB, deviant 2 differed by 1.5 ERB and deviant 3 differed by 2.25 ERB from the standard /i/. Subjects' responses in this experiment were measured in percent correct. There were three different pairs of stimuli in this experiment: pair 1 (standard /i/ vs 0.75 ERB deviant /i/; mean percent correct=39, SD=17), pair 2 (standard /i/ vs 1.5 ERB

deviant /i/; mean percent correct=52, SD=19) and pair 3 (standard /i/ vs 2.25 ERB deviant /i/; mean percent correct=69, SD=22). Based on these results from the pilot we decided to use the stimuli with 1.5 ERB deviance from the standard for our experiment as the other ERB values were either too easy or to difficult to perceive.

The stimuli for the MEG experiment consisted of four different synthesised /b/-V-/t/ realisations and were based on stimuli from a previous study (Iverson & Evans, 2007). They were designed to model a recording of a male British English speaker (the /b/ burst and the /t/ release were cut from a natural recording of this speaker). The duration of each word was 464 ms (260 ms for the vowel, excluding the bursts and /t/ stop gap).There were two /bit/ realisations (the standard stimulus: F1=210Hz, F2=2707Hz and Deviant 1: F1=245Hz, F2=2492Hz) one /bɪt/ realisation (Deviant 2: F1=372Hz, F2=2230Hz) and one /but/ realisation (Deviant 3: F1=261Hz, F2=1197Hz). As before, vowel duration was equated, F3 remained constant (F3=3200Hz) and there was no formant movement. Deviant 1 was designed to be a within-category deviant of the standard /i/ stimulus, Deviants 2 - /ɪ/ - and 3 - /u/ - were both between-category deviants for L1 speakers.

Procedure

Behavioural measures

Subjects performed an auditory discrimination (AD) task which used an oddity paradigm. The subjects heard three synthetic b-V-t combinations (which were within-category deviants of /bit/) immediately after one another. ISI was 20ms. Two items were the same, one was different (for example /bit/, /bit/ and a deviant /bit/). The subjects chose the different word by the appropriate number on the computer monitor. The experiment consisted of 6 different items (3 x standard /i/, 3 x variant /i/) which were repeated 16 times, making 96 trials in total. Every 25 trials there was an optional break. Stimuli were presented in random order. Subjects started with a short warm-up period (five items) in order to become familiarised with the task. Results from the warm-up periods were discarded for further analysis.

A vowel identification (ID) task was performed by the subjects: they heard either "beat", "bit" or "boot" and chose the appropriate word on the monitor. Each of the five female and five male speakers contributed 3 different realisations of "beat", "bit" and "boot" (90 items). The experiment consisted of 180 items in total, 90 items had normal vowel duration, 90 items were identical items which had been equalised as to vowel duration. Every 30 trials there was an optional break. Stimuli were presented in

random order. Subjects started with a short warm-up period (eight items) in order to become familiarised with the task and the stimuli. Results from the warm-up periods were discarded for further analysis.

An oddity paradigm was used for the category discrimination (CD) task. The subjects heard three words immediately after one another. ISI was 20ms. Two words were the same (but pronounced by different speakers) and one was different (for example "bit", "boot", "bit"). The subjects chose the different word by the appropriate number on the computer monitor. The experiment consisted of 48 different items which were repeated four times, making 192 trials in total. Every 30 trials there was an optional break. Stimuli were presented in random order. Subjects started with a short warm-up period (12 items) in order to become familiarised with the task. Results from the warm-up periods were discarded for further analysis.

Production was assessed by a recording of each subject reading a short story in English which had been written to include ten items containing the English phoneme /i/ and ten items containing the English phoneme /ɪ/ in a stressed position. Subjects read the story to themselves first in order to become familiarised with the content and then aloud while they were being recorded.

MEG

A VSMMedTech Omega 275 Magnetoencephalography (MEG) scanner was used to measure the electromagnetic field changes that occurred during the experiment from 274 SQUIDS (each referenced to a third order axial gradiometer) arranged around the head. 600 data points were sampled each second with an anti-alias filter applied at 120Hz. Auditory stimuli were presented binaurally using E-A-RTONE 3A audiometric insert earphones (Etymotic Research, Inc: Illinois, USA) that were attached to the rear of the subject's chair and connected to the subject using flexible plastic tubing. The stimuli were presented at 60 dB/SPL. The subjects were given the opportunity to change the stimulus loudness but this was declined by all subjects.

A passive odd-ball paradigm was used involving the auditory presentation (SOA=1080ms, ISI=0ms) of a train of repeating standards interleaved in a pseudo-randomized manner with presentations of deviant 1 (D1), deviant 2 (D2) or deviant 3 (D3). This oddball paradigm was chosen because it was one of several experiments (Leff et al., 2009; Schofield et al., 2009; Teki et al., 2013) which were run with the same stimuli, or subsets of these, and using the same mismatch oddball paradigm, however with different groups of subjects (aphasics, native speakers and second language speakers). This way all experimental data could be pre-

processed, modeled and analysed with very similar code which saved time and make the results of the experiments somewhat comparable.

Stimuli were 464 ms in duration (vowel duration of 260 ms). Within each acquisition block 30 deviants (of each sort) were presented to create a standards to deviants ratio of 4:1. A minimum of two and a maximum of six standards were presented between deviants. A total of 4 acquisition blocks were performed by each subject, resulting in a total of 120 trials for each of the 3 deviants. Each block lasted for 540 seconds. Each subject performed 480 trials/deviant with 4 blocks in total, resulting in 1440 deviant trials/subject.

During stimulus presentation subjects were asked to complete an incidental visual detection task and to ignore the auditory stimuli. Subjects were requested to attend to the visual modality to ensure their attention was focused on the pictures instead of the sounds they were hearing. However, this request may have been superfluous as the visual task demanded their full attention.

Static pictures of outdoor scenes (44 per acquisition block) were presented for randomised lengths of time, interrupted by presentation of a picture (presented for 0.5 seconds) of a circle. The subjects were asked to press a response button (right index finger) for the circles. This task was designed to ensure subjects were

attending to the visual modality.

The tasks were completed in the same order for each participant: First the MEG data was collected, then the behavioural experiments AD, ID, CD and the production task were performed. The MEG task was always performed first, at a time when the participants did not know that the experiment entailed English and German vowel perception. During a pilot, I realised that performing the behavioural tasks alerted the participants to the fact that I was looking at specific vowels and a comparison between English and German. I was worried this knowledge might have an unintended effect on their performance during the MEG task, so all participants performed the MEG task first. Due to this strategy, I hoped that they would be attentive to the distractor task and not to the auditory stimuli.

Pre-processing

Pre-processing the raw data consisted of several steps. MEG data were analysed as event-related potentials. SPM8 software (Wellcome Trust Centre for Neuroimaging, London, UK), running under Matlab 2010b (Mathworks Inc., Sherborn: USA) was used for statistical parametric mapping (Litvak et al., 2011). Statistical analysis was performed in sensor space.

For each subject, the electromagnetic field data from all

channels for each acquisition block were digitally filtered with a third-order Butterworth band-pass of 1-20 Hz. Epochs that contained a value greater than +/- 3000 femtotesla were rejected as eyeblink artifacts and removed from further analysis. Data were then organized into epochs referenced to stimulus presentation, running from 100ms pre-stimulus to 400ms post-stimulus. Epoching separates specified chunks of the data stream into epochs/trials. An epoch/trial is defined as a certain time chunk around a stimulus onset.

The mean amplitude of the pre-stimulus interval was used as a baseline for the post-stimulus data-points. Baseline correction is defined as subtracting the mean of the pre-stimulus time from the whole trial. This is done in order to obtain a period of time which is thought to not be influenced by stimulus-induced brain activity. Subtracting the pre-stimulus time period from the overall trial is thought to result in the brain activity which relates to the stimulus alone and not to other, non stimulus-related (for example due to muscle movement, boredom, fatigue etc.), responses.

Data was low-pass filtered again. Single trials within trial type are merged and then averaged (Litvak et al., 2011) in order to be able to see an elicited brain response. Data was low-pass filtered again after robust averaging to remove any high-frequency noise generated by the robust averaging procedure.

Analysis

One image per trial and multiple images per condition were generated from root-mean-square values, combining the two planar gradiometers at each location. This process generates one image per trial and multiple images per condition.

A 3D image was constructed for each epoch, electromagnetic fields being interpolated between sensors across 2D sensor space at each sample point in time. The data at each sensor at each time point were projected onto a 2D scalp map. This interpolated scalp map was a 64x64 grid. These spatiotemporal images were entered into an SPM analysis, for each subject, with four levels of one factor (STD, D1, D2, D3). The search space included all voxels in sensor space between the time-points 150-250 ms post-stimulus. The data were examined with F-contrasts of the form [-1 1 0 0] to identify the MMF for D1, [-1 0 1 0] to identify the MMF for D2 and [-1 0 0 1] to identify the MMF for D3. The value of the contrast estimate at the peak voxel in the resulting F-maps and the peak latency at this voxel were extracted for each contrast to provide summary statistics for each contrast/deviant and subject.

In order to identify MMFs in source space in an unbiased manner, the individual M100 was fitted to the standard condition (after visual inspection of the butterfly plots by researchers Leff and Oliver independently) and then the MMF was fitted to the individual

M100 for each subject. This individual fitting process was used because of the variance in latency of the M100 between individuals as fitting the MMF to the M100 accommodated individual differences between subjects which could be due to anatomical differences. The average latencies of the M100 in the L1 and L2 speakers were 103±22ms and 99±23ms.

The mismatch field for each hemisphere was computed by identifying the next positive- or negative-going peak after the M100 within a specified time-window (150-250ms post-stimulus). Due to individual differences in the latency of the M100 the difference waves and computed MMFs for each subject were visually inspected by two independent judges with no knowledge of the identity of the subject and were identified as being in the time-window 100-300ms post-stimulus. Mismatch peaks for all three deviant-types per subject were specified as being necessarily uniformly negative- or positive-going.

Results

Behavioural results

The analyses of the behavioural tasks were conducted on arcsine-transformed scores. As expected, there was no main effect of native language on auditory discrimination (AD) ($t=-.334$, $df=20$, $p>.735$).

This indicates that very similar auditory discrimination abilities in both language groups exist (see Figure 1). Therefore, we expect that all other measured differences would not be due to differences in hearing abilities between subjects.

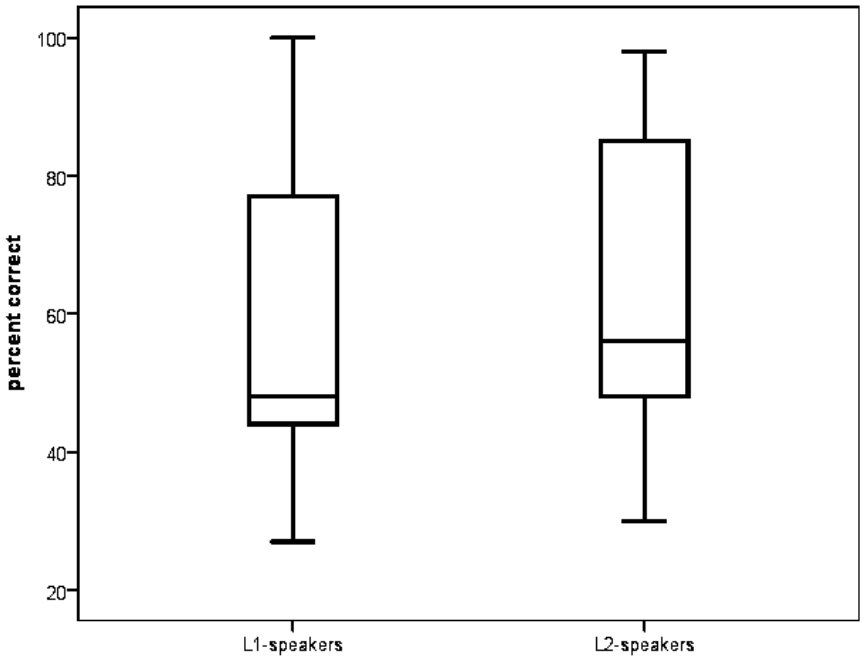

Figure 1: Percent correct score AD for L1 and L2 speakers

Subjects' responses in the vowel identification (ID) experiment were measured in percent correct. Paired-sample t-tests on the mean percent correct response for normal duration vs. equated duration vowels on arcsine-transformed scores were performed and showed a significant difference between the two types of stimuli (t=3.51, df=8, p=.008) for the L1 speakers, but not for the L2 speakers (t=-1.11, df=12, p=.290), see Figure 2.

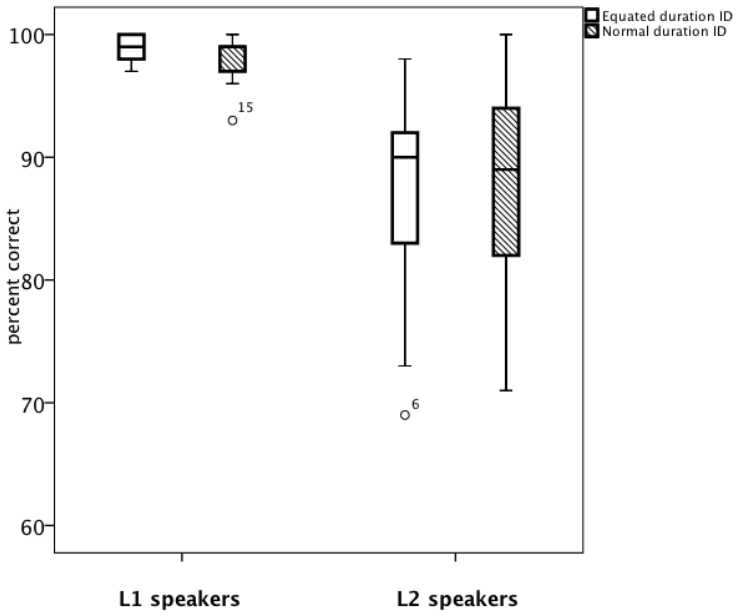

Figure 2: Normal vs. equated vowel duration for L1 and L2 speakers

The highly correlated relationship between the two conditions is visible in Figure 3 (r=.968, p<.01).

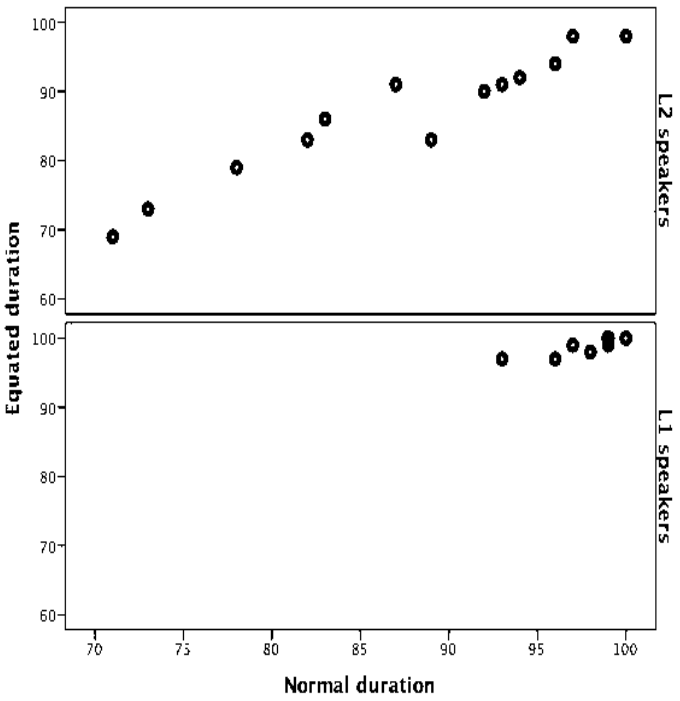

Figure 3: Normal vs. equated vowel duration for L1 and L2 speakers (percent correct)

The subjects showed cross-language differences in the vowel identification responses (see Figure 4).

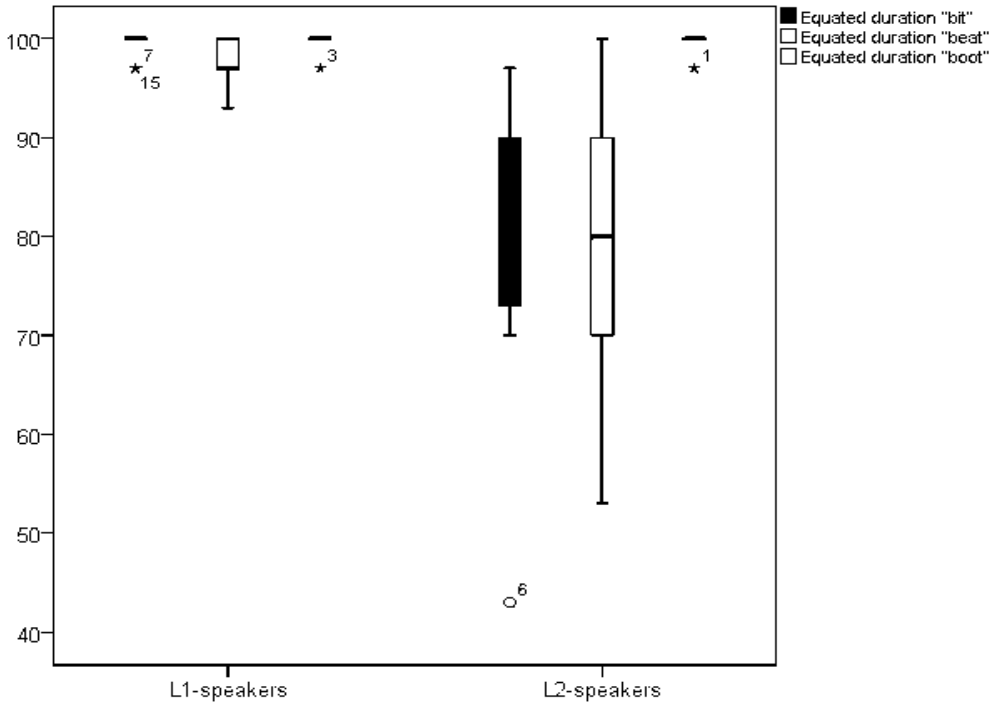

Figure 4: Boxplot of results for vowel identification task (percent correct). Equated duration vowels for L1 and L2 speakers

L2 speakers performed worse than L1 speakers, as expected, and showed a great deal of variability in their equated duration vowel ID scores. While there were large cross-language differences on ID for /i/ (t=3.75, df=20, p=.002) and for /ɪ/ (t=3.67, df=20, p=.006), there was no cross-language difference for /u/ (t=-.26, df=20, p=.61). We used only the equated duration stimuli for further analysis as we used equated duration stimuli for the MEG experiment. This makes the results of both experiments more

comparable.

Subjects' responses in the category discrimination experiment were measured in percent correct. A paired-sample t-test for the L1 speakers on the mean percent correct response for the category discrimination pair /i/ and /ɪ/ (mean percent correct=93.8, SD=5.7) vs. the category discrimination pair /i/ and /u/ (mean percent correct=93.8, SD=3.7) was performed (t=.75, df=8, p>.475) and showed the ability of L1 speakers to discriminate between all three types of stimuli equally.

A paired-sample t-test on the mean percent correct response for the discrimination pair /i/ and /ɪ/ (mean percent correct=65.4, SD=15.9) vs. the discrimination pair /i/ and /u/ (mean percent correct=93, SD=5) was also performed for the L2 speakers and showed a significant difference in the ability of subjects to discriminate between all three types of stimuli (t=-7.564, df=20, p>.001). Cross-language differences in the category discrimination responses were observable. There was an effect of native language on CD /i-ɪ/ (see Figure 5) (t=6.017, df=20, p<0.00). L2 speakers performed worse than L1 speakers in this respect, as expected.

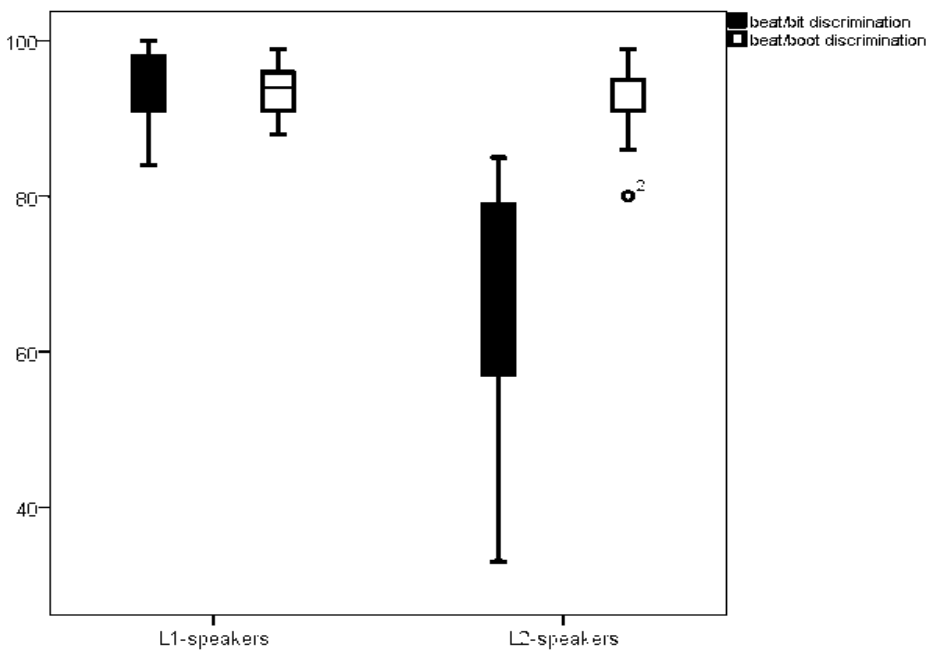

Figure 5: Boxplot of category discrimination values between the vowels /i/ and /ɪ / and between the vowels /i/ and /u/ in percent correct, grouped by native language

For the production task we looked at the acoustic contrast between the target phonemes /i-ɪ/ in the L1 and the L2 speakers' production. The spectral difference between /i-ɪ/ was calculated, based on Euclidean distance combining F1 and F2. F1 and F2 for each target phoneme were measured, using PRAAT. Mean distance for L1 speakers between F1 and F2 was 286.2 Hz, SD=112.7. Mean distance for L2 speakers between F1 and F2 was 89.1 Hz, SD=98. There was a wide range of results, the distance varying from 109 to 413 Hz between F1 and F2 for L1 speakers and from 11 to 324 Hz for L2 speakers (see Figure 6).

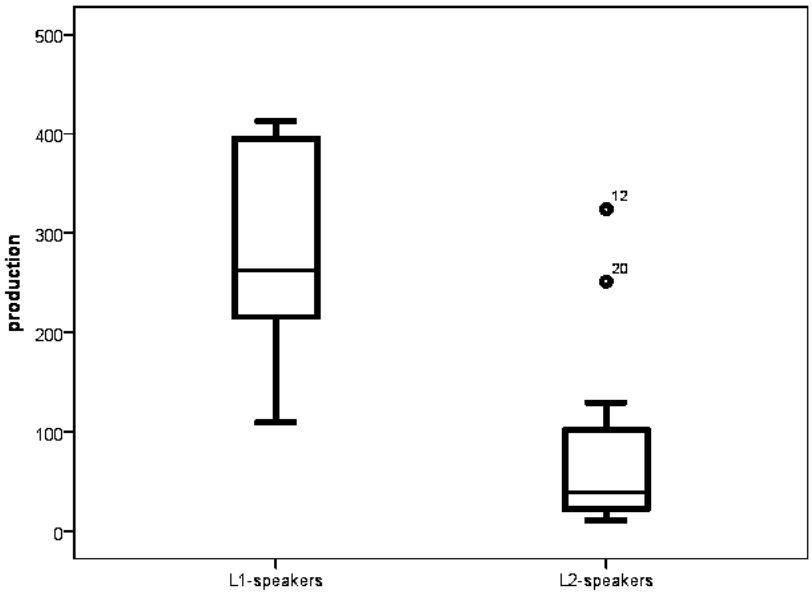

Figure 6: Vowel production L1 and L2 speakers (in Hz)

As expected, cross-language differences in the production results were visible. There was an effect of native language on production in that L2 speakers performed worse than L1 speakers (t=39.7, df=21, p<.0001). L2 speakers were less able to differentiate between /i/ and /ɪ/ than L1 speakers. For L1 speakers, only ID beat and ID bit (r=.876, p=.002) and, additionally, AD and CD beat/boot (r=.800, p=.01) correlated highly with each other (see Table 1). Production did not correlate with any other scores for the L1 speakers. For L2 speakers, production and ID beat scores correlated moderately with each other (r=.573, p=.04) and CD beat/bit and CD beat/boot scores correlated moderately with each other (r=.647, p=.017), while AD did not correlate with any other scores (see Table 2).

N=9	production	AD	ID beat	ID bit	ID boot	CD beat/bit	CD beat/boot
production	1	.277	.652	.516	.079	.321	.201
AD	.277	1	.598	.404	.137	.467	.800*
ID beat	.652	.598	1	.876*	.058	.218	.598
ID bit	.516	.404	.876*	1	-.189	.323	.644
ID boot	.079	.137	.058	-.189	1	-.405	-.420
CD beat/bit	.321	.467	.218	.323	-.405	1	.631
CD beat/boot	.201	.800*	.598	.644	-.420	.631	1

Table 1: Correlation matrix of behavioural measures for L1 speakers

N=13	production	AD	ID beat	ID bit	ID boot	CD beat/bit	CD beat/boot
production	1	-.444	.573*	.424	-.017	.522	.355
AD	-.444	1	.036	-.267	.180	.192	.216
ID beat	.573*	.036	1	.700*	.548	.432	.293
ID bit	.424	-.267	.700	1	.231	.118	-.192
ID boot	-.017	.180	.548	.231	1	-.371	-.118
CD beat/bit	.522	.192	.432	.118	-.371	1	.647*
CD beat/boot	.355	.216	.293	-.192	-.118	.647*	1

Table 2: Correlation matrix of behavioural measures for L2 speakers

Paired-samples t-tests on the results were performed which showed that while there was no significant differential effect of language upon auditory discrimination, there was however a significant effect of language on category discrimination for /i-ɪ/ ($t=33.68$, $df=21$, $p<.000$), on vowel identification ($t=29.31$, $df=21$, $p<.000$) and on production ($t=39.77$, $df=21$, $p<.000$).

MEG results

Each deviant stimulus was presented 120 times per block, with 4 blocks in total, which made 10560 trials in total per deviant for 22 subjects (4 blocks x 120 trials = 480 trials/deviant; 480 trials x 3 deviants = 1440 trials/subject; 1440 trials x 22 subjects = 31680

trials total for all subjects). 31680 trials in total for all blocks and deviants from all 22 subjects (L1 and L2 speakers) were pre-processed for analysis, 1997 trials from all three deviants (667 trials from D1, 661 trials from D2 and 669 trials from D3) were rejected due to eyeblinks or other artefacts, this being 6.3% of the total trials. Of the remaining 29683 deviant trials, 9893 were D1, 9899 were D2 and 9891 were D3 trials and were used for further analysis. In total, there were 12960 trials for L1 speakers (120x4x9) and 18720 trials for L2 speakers (12x4x13) for all three deviants which were used for further analysis.

The results from the MEG analysis indicated that there was no general effect of language on MMF peak intensity or latency with our subjects, in that L1 speakers did not perform differently than L2 speakers on the MEG task. Therefore, Figure 7 shows the MMF for both L1 and L2 speakers together at sensor location MLT14, which is known to be above the auditory cortex (Schofield et al., 2009), illustrated by a butterfly plot. Although analysis was performed over all electrodes I chose sensor MLT14 for illustrative purposes, as a graph averaged over all sensors would not show the typical MMF peaks.

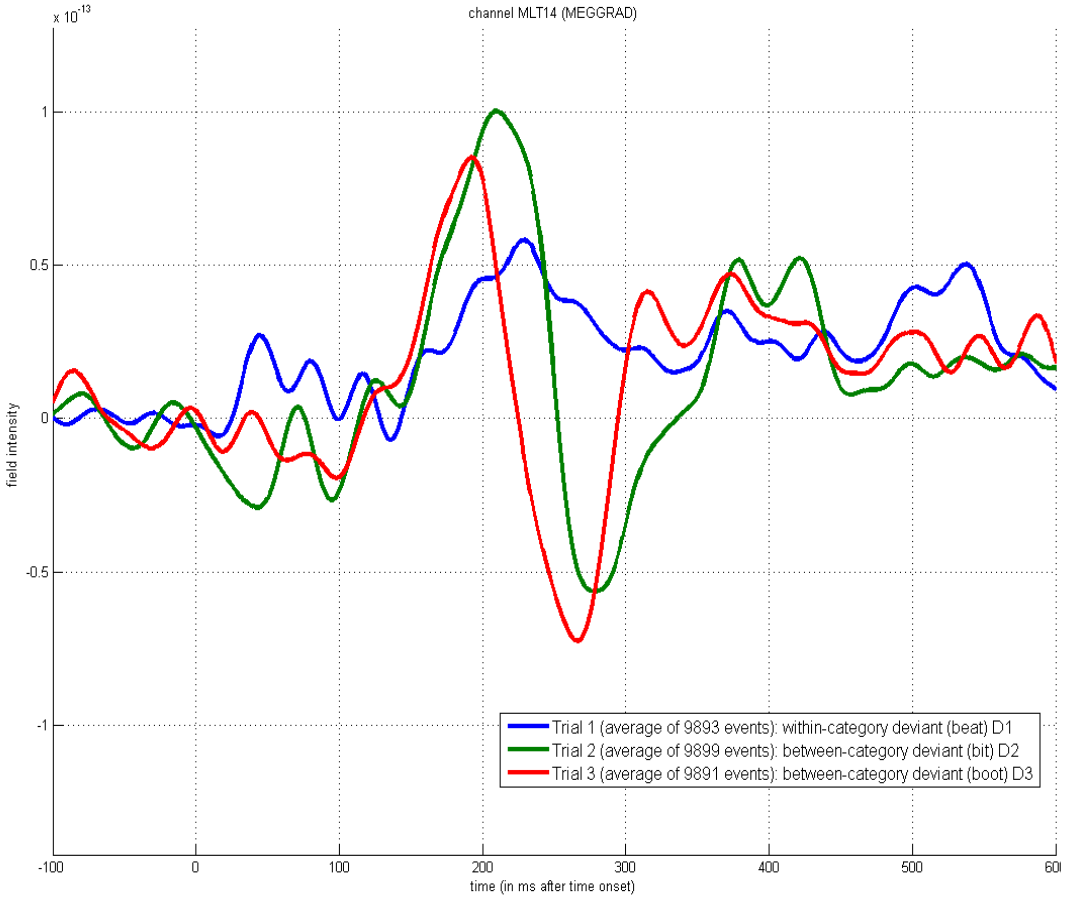

Figure 7: MMF for all subjects, L1 and L2 speakers, for all deviants, D1 (blue), D2 (green) and D3 (red) at sensor location MLT14. Each deviant shows a differentiated mismatch response.

However, there was an effect of language experience on performance in category discrimination, ID and production, in that L2 speakers performed at a lower level than L1 speakers in all these tasks.

Values for amplitude and latency are shown split by hemisphere (left versus right) for L1 and L2 speakers in Figures 8 and 9. Hemispheres were included as factors in the analysis in order to investigate whether there were indeed hemispheric differences in

general, as posited in the literature so far (Friston et al., 2003; Luo et al., 2006; Naatanen et al., 2007; Näätänen et al., 1997; Okamoto & Kakigi, 2013; Teamu Rinne et al., 1999; Tervaniemi & Hugdahl, 2003) and also whether there are differences between L1 and L2 speakers with respect to hemispheric dominancy or differences.

As expected, a left-hemisphere dominancy of the MMF was visible, indicating that the brain recognized the stimuli as speech-sounds. Different MMFs for each deviant type were elicited, indicating that the stimuli and experimental setup were adequate for our purposes.

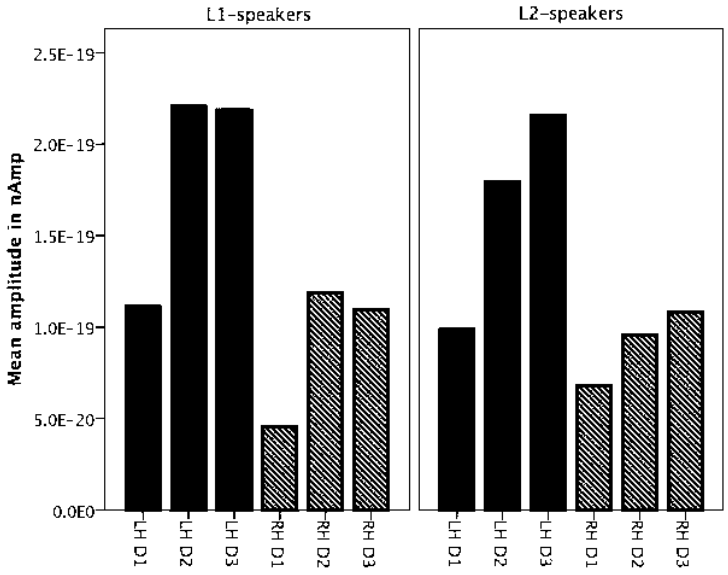

Figure 8: MMF amplitude in nAmp by group (L1 and L2 speakers)

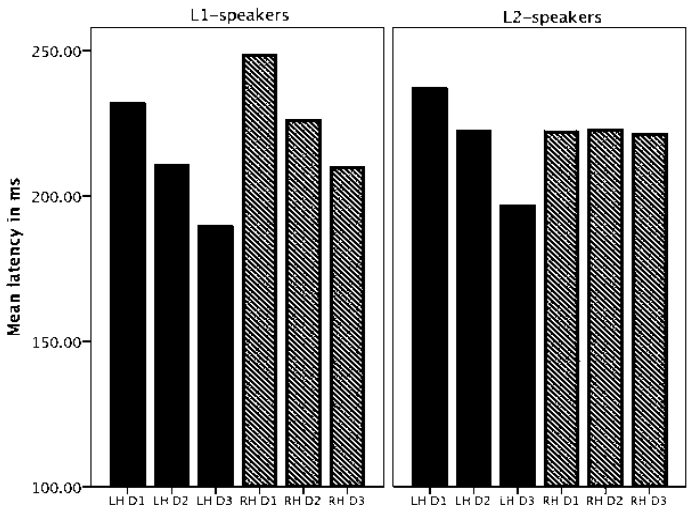

Figure 9: MMF latency in ms by group (L1 and L2 speakers)

Analysis on raw scores

In order to see which measures had an influence upon the amplitude of the mismatch response, the MMF amplitudes for both L1 speakers and L2 speakers were entered into a mixed model analysis with subject as a random factor and the fixed factors deviant (D1, D2 and D3), laterality (LH/RH), language (L1/L2 speaker), auditory discrimination, category discrimination /i-ɪ/, equated duration vowel identification and production. This analysis yielded significant main effects of deviant ($F(2,80)=3.92$, $p=.02$) and laterality ($F(1,80)=10.04$, $p=.002$). Additionally, analysis showed that there were neither two- nor three-way significant interactions between any of the factors. This mixed model analysis indicates that different deviants show differentiated MMF peak amplitudes and that the location of the MMF (left- versus right-

hemisphere) has an influence upon the height of the MMF peaks in that MMF peaks for the right hemisphere generally are lower for all subjects.

In order to see which variables had an effect upon the latency of the mismatch response, MMF latencies for L1 speakers and L2 speakers likewise were entered into a mixed model analysis with subject as a random factor and the fixed factors deviant, laterality, language, auditory discrimination, category discrimination /i-ɪ/, equated duration vowel identification and production. The analysis yielded a significant main effect of deviant ($F(2,80)=10.2$, $p<.001$). Additionally, this analysis showed that there was a significant interaction between the factors laterality and production ($F(1,80)=4.4$, $p=.03$), see Figure 10 and Figure 11. The mixed model analysis indicates that each type of deviant produces a different MMF latency and that MMF latency and production results vary, depending on the hemisphere. This is probably only a relatively small effect and we have no explanation for it.

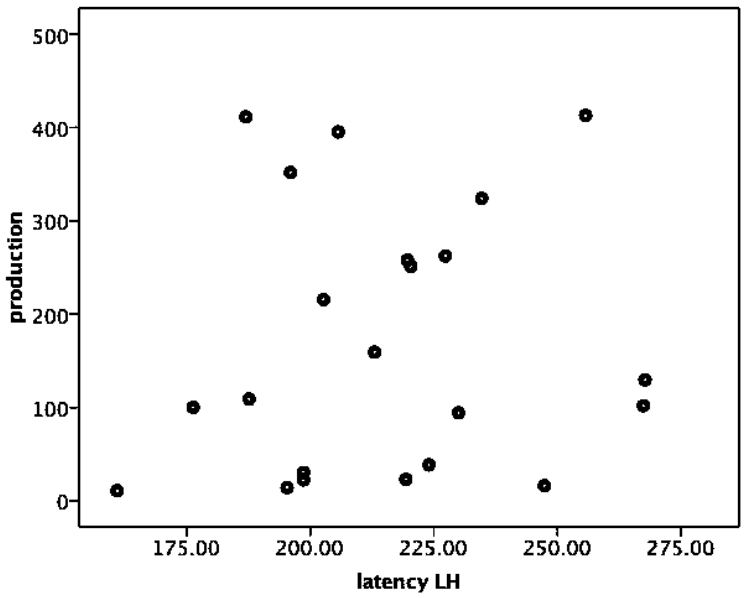

Figure 10: Production/latency interaction for the left hemisphere for all subjects

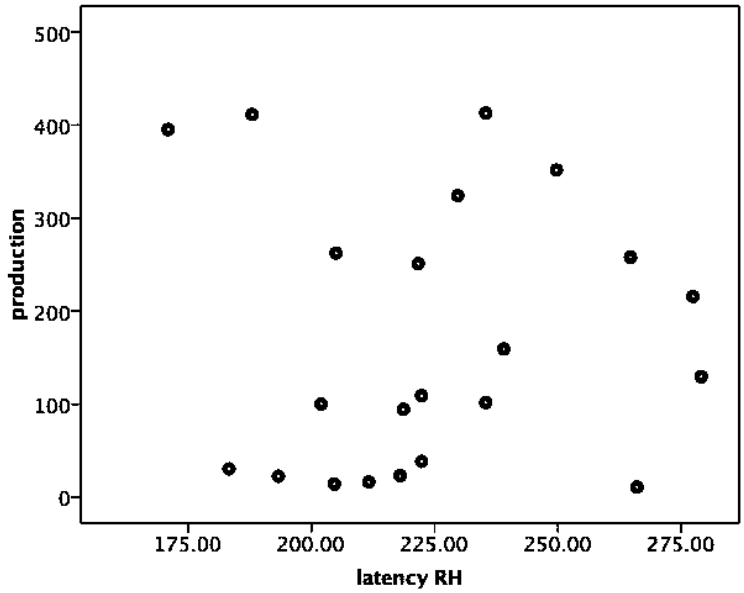

Figure 11: Production/latency interaction for the right hemisphere for all subjects

The mixed model analysis indicates that each type of deviant produces a different MMF latency and that MMF latency results vary, depending on the hemisphere.

In order to further examine the effect of language background

upon the brain responses, we performed a partial correlation analysis which controlled for the influence of variables which were not directly of interest to our research question. In our experimental design, deviants 1 and 2 were conceived as controls for the experimental variable D2, which investigates whether subjects can hear a difference between a phoneme that exists in their native language and one which does not. Deviant 1 confirms that all subjects can perceive acoustic deviance from the standard phoneme, regardless of language background. Deviant 3 confirms that all subjects can perceive differences between two different phonemes that exist in their own languages, also regardless of language background.

Deviants 1, 2 and 3 for amplitude for each hemisphere are very highly correlated with each other (see Tables 2 and 3) for all subjects (L1 and L2 speakers), this correlation is probably driven by brain structure similarity within subjects which influences all MMF data to a large extent.

Left hemisphere	Deviant 1	Deviant 2	Deviant 3
Deviant 1	1	.902**	.802**
Deviant 2	.902**	1	.889**
Deviant 3	.802**	.889**	1

Table 3: Correlations between deviants for left hemisphere (MMF amplitude) for L1 and L2 speakers

Right hemisphere	Deviant 1	Deviant 2	Deviant 3
Deviant 1	1	.846**	.843**
Deviant 2	.846**	1	.743**
Deviant 3	.843**	.743**	1

Table 4: Correlations between deviants for right hemisphere (MMF amplitude) for L1 and L2 speakers

In order to partial out the effects of D1 and D3 on the data and only look at the correlation of D2 (our variable of interest) with the behavioural factors, we controlled for D1 and D3 while performing partial correlations of D2 amplitude and latency with the behavioural data for L1 and L2 speakers. We concentrated on the left hemisphere as this is seen to be the hemisphere which is responsible for auditory speech processing. A partial correlation analysis of left-hemisphere D2 amplitude (with the factors LH D1 amplitude and LH D3 amplitude controlled for) with language, auditory discrimination, category discrimination for /i-ɪ/, equated duration vowel identification and production was performed which showed a significant effect of category discrimination for /i-ɪ/ ($F(1,80)=5.66$, $p=.05$) on LH D2. A partial correlation analysis of left-hemisphere D2 latency (with the factors LH D1 latency and LH D3 latency controlled for) with language, auditory discrimination, category discrimination for /i-ɪ/, equated duration vowel identification and production was performed which showed no

significant effects of any of our factors.

Discussion

Our study indicates that there are cross-language differences in the behavioural responses. Auditory discrimination, category discrimination, and identification all involve making a post-perceptual decision and thus can involve several levels of processing. Performance on the auditory discrimination task requires listeners to discern differences in the acoustic-phonetic information; listeners may also covertly label the stimuli in terms of phonetic categories in order to reduce memory load, but this is relatively unlikely within a low-variability task using vowels. Identification involves low-level processing of the acoustic-phonetic information, and explicitly requires the listener to label the stimulus in terms of their long-term representations for these categories. Identification depends on lower-level discrimination processes such as auditory discrimination. Category discrimination explicitly requires listeners to pay attention to the phonetic information that distinguishes categories, but it is the least pure of the three measures in terms of tapping a particular level. The task does not require phoneme labelling, but covert labelling might be useful given the relatively high stimulus variability, at least for listeners

with well developed category representations. Also, listeners need to ignore the acoustic variation that is irrelevant to phonetic categorisation when doing CD, but having sensitivity to the relevant phonetic dimensions is important too.

There was an effect of native language on CD for /i-ɪ/, ID and production. L2 speakers performed worse than L1 speakers in all the behavioural measures – this result was as expected. There was no main effect of native language on AD which indicated that very similar auditory discrimination abilities in both language groups existed; the other measured differences can therefore be thought not to be due to a differing ability to hear or discriminate between sounds acoustically. L2 production and perception were only moderately correlated; this is thought to be due to perception and production drawing on different abilities, so that even in highly proficient L2 subjects these abilities are not highly correlated. Previous training studies indicate that it is probable that production and perception are only moderately correlated and driven by different factors or even underlying representations (Hattori & Iverson, 2009).

Motor and gestural theories posit a very strong dependence of perception on production processes and representations and links between sensory input and motor speech systems (Galantucci, Fowler, & Turvey, 2006; Lane, 1965; A. M. Liberman, Cooper,

Shankweiler, & Studdert-Kennedy, 1967; A. Liberman & Mattingly, 1985; A. Liberman & Whalen, 2000). Despite studies showing that hearing speech activates vocal tract muscles (Fadiga, Craighero, Buccino, & Rizzolatti, 2002), the motor cortex (Watkins, Strafella, & Paus, 2003) and premotor cortex (Wilson, Saygin, Sereno, & Iacoboni, 2004), our research has not shown the strong links between production and perception behaviourally that one would expect from motor and gestural theories. However, when observing that speech perception functions relatively well even with non-normal motor speech systems (ie. damaged, deactivated or underdeveloped systems) it seems that the influence of the motor system on perceptual abilities may be relatively limited (Bishop, Brown, & Robson, 1990; Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Hickok, Costanzo, Capasso, & Miceli, 2011; Hickok et al., 2008; Kuhl & Miller, 1975; Rogalsky, Love, Driscoll, Anderson, & Hickok, 2011; Rogalsky, Pitz, Hillis, & Hickok, 2008).

The results showed a left-hemisphere dominancy of the MMF which indicates that, as expected, the brain recognized the stimuli used as speech-sounds as compared to non-speech sounds (Näätänen, 2001). Different MMFs for each deviant type were elicited, showing that the brain distinguished between each type of deviant. There was no main effect of language on the MMF. This was probably partly due to the fact that the subjects participating in

this study tended to be highly proficient L2 speakers – with more low proficiency L2 speakers the results might look different and produce a main effect of language in the MMF (in addition to in the behavioural tasks).

Our subjects are in effect one large group who vary in their ability to perceive or be sensitive to the English phoneme /ɪ/. The ability to perceive this English phoneme could be seen as a continuum, ranging from no/low proficiency/sensitivity to high proficiency/sensitivity. The English L1 speakers should all be clustered around the top end of proficiency. However, the English L2 speakers can theoretically range all over the continuum from low/no proficiency to high proficiency. All three deviants are acoustically different from the standard phoneme /i/. As such, they should all three elicit a mismatch response as this acoustic deviation from the standard will be detected automatically by the brain. The three deviants used in our study should however elicit differentiated mismatch responses, depending on where on this proficiency continuum the speakers are located and according to the deviant type. In order to elicit a mismatch response for D1, subjects must detect the acoustic difference present between the standard /i/ and D1. This MMF elicited by D1 is language independent, as the phoneme exists in both English and French and there are merely slight acoustic differences between D1 and the standard /i/. In

order to elicit a mismatch response for D3, subjects must detect the acoustic difference between the standard and D3 and perceive that it is a different phoneme /u/ to the standard /i/. This MMF elicited by D3 is also language independent, as this phoneme exists in both English and French. D2, however, should elicit different mismatch responses depending on the subjects' ability to perceive /ɪ/.

As mentioned above, however, our population sample contained highly proficient L2 speakers and only highly proficient L1 speakers, therefore there was no language effect on the D2 MMF results. All the L2 speakers were so (relatively) highly proficient at/sensitive to perceiving the non-native phoneme (as indicated by the similar mismatch response to D2) that there was no difference between L1 and L2 speakers.

The present study found a partial correlation between category discrimination measures for /i-ɪ/ and the LH MMF amplitude for D2 /ɪ/. Vowel identification is seen as tapping into phonological categorization processes, auditory discrimination is thought to reflect the detection of acoustic differences and category discrimination is seen to be representative of phonetic/acoustic processing. According to Näätänen (Näätänen et al., 1997), the MMF represents phonetic/phonemic categorization processes. However, it is not particularly surprising that most of the behavioural tasks were only mildly correlated with each other. This

points towards the different behavioural tasks tapping into different abilities or processes and is borne out by the literature (Bradlow et al., 1997; Hattori, 2009; Iverson et al., 2011).

Studies involving auditory training with L2 speakers (Bradlow et al., 1997; Iverson et al., 2011) have shown that, while similar behavioural task results (such as identification, discrimination, perceptual best exemplar mapping and production) have correlated pre-training, post-training improvements have not correlated with each other. This implies that the training has had a different effect on different L2 perception abilities, leading to non-correlated end results. Thus, training does not affect all phonetic measures to an equal extent. Dutch listeners, for example, were trained to improve their identification of a Finnish consonant contrast, but did not improve in parallel in their ability to discriminate between acoustic changes within that contrast (Heeren & Schouten, 2008). Merely the fact that highly proficient L2 listeners are subject to a large amount of L2 input over the years will lead to a general increase in a wide range of phonetic abilities. Auditory training will, however, target specific underlying processes and lead to differentiated and non-correlated outcomes (Iverson & Evans, 2009).

However, why do the category discrimination /i-ɪ/ scores for L1 and L2 speakers correlate specifically with the MMF of the D2 deviant /ɪ/ and not with the other deviants? As mentioned earlier,

D2 is the one deviant in the present study whose elicited mismatch response is driven by vowel perception proficiency, e.g. sensitivity to the phoneme status of a sound. It is different from auditory discrimination in that it taps into phonological/phonetic levels of encoding instead of purely acoustic features. The listener must ignore irrelevant phonetic dimensions such as talker variation and needs to focus on dimensions that differentiate categories from each other - acoustic similarity or dissimilarity alone are not enough to perform this task (Iverson et al., 2011). Category discrimination tasks demand a basic perception of the phonetic dimensions that distinguish vowels. It is probable that categorization efficiency is the main source of the MMF effect in our study.

The subjects in this study were mainly highly proficient L2 speakers. Although we had attempted to sample speakers from a wide range of proficiency levels, our subjects were in fact quite proficient in most measures and fluent L2 speakers. The MMF might be picking up on their relative sensitivity to phonemes in auditory processing and categorization which is reflected in their category discrimination score. The L2 speakers are functioning at a high level with respect to their L2 auditory processing and have had a lot of L2 input. The MMF could be seen to reflect this automatic phonetic/acoustic pre-categorial processing. In this context, we would expect to see different results if the sample contained more

low-proficiency L2 speakers. In this case, not only could we expect to see an overall language effect in the behavioural results, but we could expect language group or L2 experience to be reflected in a differential D2 MMF amplitude/latency and perhaps a correlation with their ID score, reflecting not only the low-proficiency L2 speakers' incomplete/incorrect L2 category representation (lack of a correlation with their CD score) but also their relative inefficiency in auditory processing (correlation with their ID score) compared to high proficiency L2 speakers such as in this study.

Analysis of links of the language background questionnaires with all scores and elicited MMFs indicated that there was no effect of length of exposure to English as an L2 on any measures – but this is probably due to our sample, all L2 subjects being highly proficient and most subjects having been in the UK for a long time. Additionally, there was no effect of age of acquisition on any measures – however, this could be due to the very limited range of age of acquisition between subjects (the earliest subject started the L2 with 9 years old). Acquiring subjects with a more diverse range of age of acquisition may lead to an effect of age of acquisition upon the measures. Again, these results may be investigated differently if all subjects were grouped into one large group (L1 and L2 speakers together) and length of exposure was correlated with all measures for all subjects.

In sum, we conclude that, with nine native controls and thirteen L2 speakers performing four perception and production tasks and an MEG oddball paradigm, we achieved the expected results of a differential MMF for each deviant, left-hemisphere dominancy, cross-language effects on perception and production scores and only moderate correlations between perception and production scores. Additionally, our results indicate that the category discrimination task used in this study shows the status of a speech sound in a listener's brain (ie. is the sound an L2 phoneme or is it just a variation of an L1 phoneme), as measured by the MMF for D2.

Chapter 3: Dynamic Causal Modelling

Introduction

This chapter deals with the further analysis of the MEG mismatch response results from the previous MMN chapter using Dynamic Causal Modeling/DCM (Friston et al., 2003). Therefore, subjects, stimuli, apparatus, procedure and tasks are identical to the previous chapter. The previous chapter on the Mismatch Response discussed MMF peak response data. The current chapter will further analyse the MEG mismatch response data using DCM, investigating neural network connectivity, the influence of language background on brain connectivity and the links between brain and behavioural data.

The aim of neuroimaging is to explain how certain types of stimuli create certain brain states. These brain states are visible only indirectly, in that MEG data is recorded using sensors, however, this same data could have been caused by many possible underlying neuronal source configurations and interactions (Nunez & Srinivarsan, 2006). This problem of not having a one-to-one correlation between sensor and source is called the inverse problem. Dynamic causal modeling/DCM (Buechel & Friston, 1997; McIntosh & Gonzalez-Lima, 1994) is one approach to solving the inverse problem and identifying which neuronal sources and interactions cause which sensor space output.

Neuronal modelling creates plausible models of interactions among neuronal populations at a cortical level. This is done by combining theoretical models and measured brain responses, using electromagnetic or hemodynamic neuroimaging methods (ie. MEG/EEG/fMRI). The present study uses dynamic causal modelling to explore the collected MEG data.

Dynamic causal modeling/DCM (Friston, 2005, 2009; Friston et al., 2003) explains measured data as the output of an interacting network of brain regions. This network consists of different areas in the brain which work together differently depending on what type of input they receive. Some of these brain regions get direct sensory stimulus input, some do not. Differences between evoked responses are measured under different conditions and are modelled as an adjustment of specified DCM parameters (for example cortico-cortical connections). Hypotheses about connectivity between brain regions are directly testable by using a neuronal network model (with several sources) to explain measured data (Friston, 2009). DCM uses a dynamic coupling of sources in the brain (Friston et al., 2003). Network model spatiotemporal inversion is then performed.

The neuronal model has several types of parameters: Connection strength, synaptic rate constants and propagation delays between sources. The location and orientation of the equivalent current dipoles/ECDs are the spatial parameters. The

location of equivalent current dipoles is pinpointed with the use of informed priors, meaning that the location of the ECDs can vary according to model. After inversion, model spaces are searched, leading to a Bayesian comparison of evidence for different models (Friston, 2005). The best fitting model, using Bayesian model comparison, of neuronal sources and connections is chosen to explain the data (Penny et al., 2010; Penny, Stephan, Mechelli, & Friston, 2004). The negative free-energy value of each model is used to decide which is the winning model (Friston et al., 2003; Friston & Stephan, 2007).

DCM is different from conventional approaches to neuronal modelling. This is due to the fact that it uses a generative model of measured brain responses: DCM is based upon the assumption that measured brain responses are nonlinear and dynamic and that experimental manipulations lead to changes in measured responses via changes in connection strengths within or between sources (David et al., 2006; Kiebel, Garrido, & Friston, 2007). Conventional approaches to neuronal modelling assume that interactions are linear, observed responses are driven by endogenous/intrinsic noise and that input to the brain is unknown. Additionally, conventional approaches do not use experimental manipulations. In contrast, DCM assumes that observed brain responses are driven by changes in the experimental design.

DCM aims to estimate functional connectivity between brain regions and to make inferences about this connectivity. By designing changes in the experimental context, DCM tries to determine how these connections between brain regions are influenced (Friston et al., 2003). Basically, DCM aims to construct a realistic neuronal model of how cortical regions interact dynamically. Additionally, this basic model is then supplemented with a forward model (of, for example, electromagnetic or hemodynamic measurements) of how neuronal or synaptic activity is transformed into a measured brain response. Due to this combination of neuronal and forward models effective connectivity (defined as a coupling among unobserved neuronal activity in different brain regions) between brain regions can be estimated from the measured data acquired during experiments.

The unique selling point of DCM is that it unites the spatial forward model with a neurobiologically informed temporal forward model, illustrating the connectivity between sources in the brain. This leads to a more robust source reconstruction (due to the informed constrainment of spatial parameters) and lets us make inferences about connectivity between brain regions. DCM makes it possible to infer parameters not directly visible in MEG data. Additional information over and above simple evoked responses can be gained by DCM through neurobiologically informed hypothesis

testing on connective network architectures in the brain. DCM utilises a dynamic input-state-output modelling approach with multiple inputs and outputs. According to DCM, known/experimental "inputs" evoke measured responses ("outputs") via non-observable internal brain "states". The "inputs" (conventional experimental stimuli) evoke measured responses in two different ways: either by directly influencing specific anatomical nodes (such as visual stimuli evoking a visual cortex response) or by influencing the functional connections between nodes/brain regions. The sort of experimental variables which could produce this second kind of measured response would be more permanent, such as attention to a specific feature.

A more intuitive way of visualising this type of modelling approach is to see an experiment as designed disturbance of neuronal dynamics (Friston et al., 2003). This disturbance is distributed through a system of inter-connected neurons and changes region-specific brain activity. Through these changes, experimental design-specific changes in brain responses are measured. These design-specific changes in brain responses are then used to estimate brain structure and functional connectivity between brain regions. The connectivity between brain regions is defined using extrinsic and intrinsic/self connections.

Experimental design

DCM is not used in an explorative way – it is used to test specific hypotheses by specific experimental manipulations. There are two ways in which inputs can elicit brain responses/outputs. Inputs can cause changes in brain states directly (for example: the auditory input eliciting direct responses in the primary auditory cortex) or they can cause changes in effective connectivity between brain regions (for example: the attentional modulation of connections between parietal and extrastriate areas).

Regarding the design of factors for experiments, factors can be regarded as belonging to one of two classes. Either factors can be classified as inputs that elicit evoked responses or they can be classified as inputs which are contextual and elicit changes in brain region coupling. Inputs that elicit evoked responses are trial- or stimulus-bound experimental factors while contextual inputs establish a context in which effects of the first sort evoke measured responses. For example, speech sounds belong to the first class, eliciting an immediate response from the auditory cortex. What native language a subject speaks would establish the context in which these speech sounds are then further processed in the brain. Depending on whether the heard speech sounds belong to the subject's native language or not, the stimuli are processed differently in the brain.

Based on the finite amount of information storage capacity in the brain, it has been proposed (Huang & Rao, 2011; Rao & Ballard, 1999) that early sensory processing is used to reduce information redundancy and to transfer sensory input into a more efficient form in the nervous system. The predictive coding model is the most influential model that explains increased efficiency in processing sensory input.

Predictive coding

According to the predictive coding model (Friston & Kiebel, 2009), the brain actively predicts input rather than only passively reacting to it. Prediction in this case means that the brain predicts the input it is most likely to get, depending on the context from the recent past. Predictions can be understood as statistical world knowledge or probabilities of occurences, patterns, rules and connections, learned by neural networks in the brain. Predictions (which are generated by the higher processing level areas) are compared to the actual input coming into the brain via the senses. If there is a mismatch between input and prediction, a prediction error is generated, or error signal, which results in brain activity in the lower sensory brain areas. Following this activity, the error signal gets passed up to higher level processing areas and an adjustment of predictions in the brain occurs (see Figure 12).

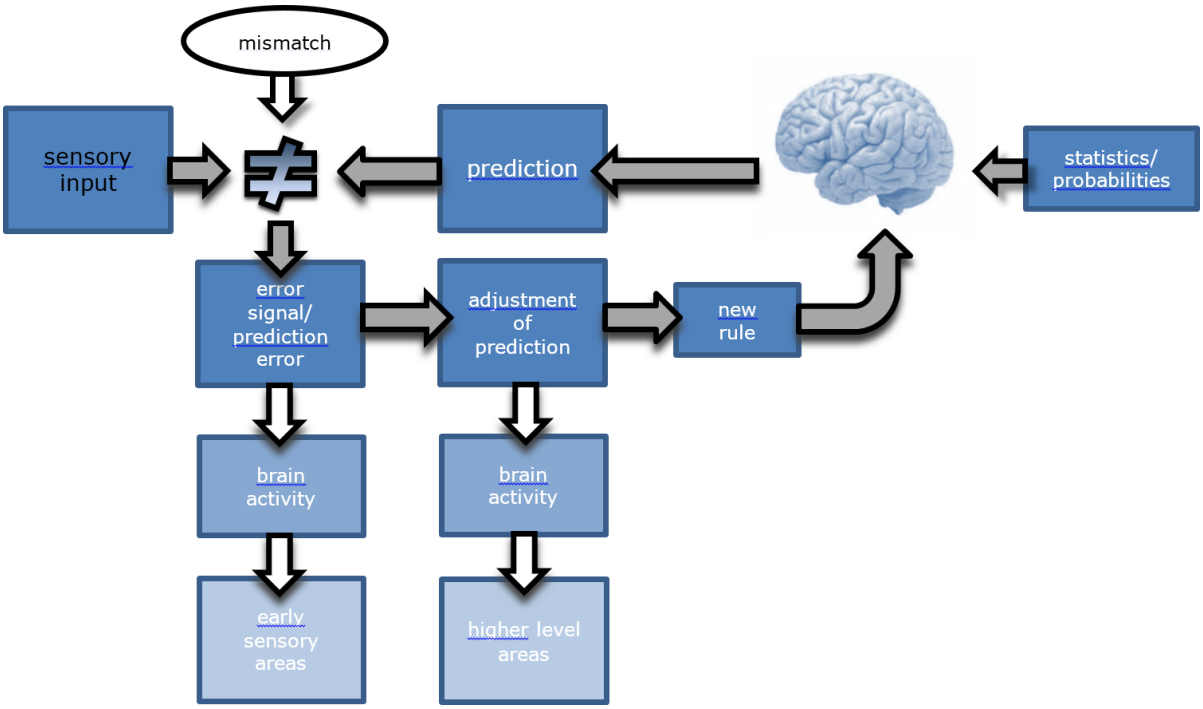

Figure 12: Predictive coding model

Prediction error can be seen as an error signal that triggers prediction adjustment in higher level processing. According to this model, brain activation present at the early sensory levels represents information not predicted by higher levels of processing and does not represent the total sensory information coming in (Rao & Ballard, 1999). This means that predictive coding offers the benefit of redundancy reduction – there is increased efficiency and automaticity as there is no need to keep multiple versions of the same information at different processing levels. Only the unpredicted parts of the incoming sensory signal are passed on to higher processing levels for further processing (Huang & Rao, 2011).

The free energy model (Friston & Kiebel, 2009) looks at how neural dynamics can be explained in terms of prediction errors. It investigates how hidden causes in a hierarchical dynamic model of the world are estimated by optimising free energy/prediction error. This model explains perceptual inference and complex cognitive phenomena. According to the free energy principle, prediction error leads to learning through synaptic plasticity. This adaptive learning is due to connection strengths being reconfigured and leads to the minimisation of prediction error at both sensory and motor levels (Friston & Stephan, 2007).

A study (Alink, Schwiedrzik, Kohler, Singer, & Muckli, 2010) concentrating on the primary visual cortex, using fMRI, indicate that there is a lower BOLD signal for predictable vs unpredictable stimuli. Another study (den Ouden, Daunizeau, Roiser, Friston, & Stephan, 2010) using visual stimuli and fMRI found that visual stimuli were differentially predicted by auditory cues when the predictability of the cues varied over time. This variation in predictability led to stimulus probabilities varying over time. Probability estimates therefore needed to be updated continuously. Behaviourally, speed and accuracy of the motor responses increased significantly with improvements in predictability. A DCM analysis of the data demonstrated that striatal prediction errors were used to improve functional coupling in cortical networks during learning. This study

indicates that the degree of prediction error activity controls the efficacy of visuomotor connections and the influence of surprising stimuli on premotor activity, showing evidence for prediction error-dependent plasticity. These studies (Alink, Schwiedrzik, Kohler, Singer, & Muckli, 2010; den Ouden et al., 2010) indicate that the brain is sensitive to predictability and that higher levels of predictability lead to a reduction of activation in sensory areas.

DCM and the Mismatch Response

It is thought that the MMN reflects the brain updating a model of the acoustic environment, in that the brain firstly registers an incoming sound and secondly makes a comparison with an existing model of expected sounds in the acoustic environment and a subsequent update of this acoustic model if it proves incompatible with what is heard (Sussman & Winkler, 2001; Winkler, Karmos, & Näätänen, 1996a). Evidence (T. Rinne, Alho, Ilmoniemi, Virtanen, & Näätänen, 2000) suggests that the frontal and temporal MMN sources behave distinctly over time, both sources seeming to interact with each other (Jemel, Achenbach, Müller, Röpcke, & Oades, 2002).

In our work, we tested the hypothesis that different abilities to categorise a speech sound are reflected in different connections and connection strengths in a neuronal network model. We motivated our prior source locations using findings from the fMRI,

MEG and EEG mismatch response literature (Garrido et al., 2008; Garrido et al., 2007; Schofield et al., 2009; Schofield et al., 2012; Teki et al., 2013). We postulated four sources over left and right primary auditory cortices (A1) and left and right superior temporal gyri (LSTG and RSTG). Left and right primary auditory cortex (LA1 and RA1) served as cortical input stations for the auditory speech sound stimuli. A1 were connected to ipsilateral STG, inter-hemispheric connections were placed between STGs and A1s, and all connections were reciprocal (forward and backward connections). We tested 255 models using this network structure – all possible forward and backward connections between left A1, right A1, left STG and right STG and self-connections for all four sources (apart from cross-connections). All models were compared against a baseline model, which had the same sources, but no connections between sources.

MEG source space analysis

MEG source reconstruction is a complex procedure due to the inverse problem: Estimating the neuronal sources from the MEG scalp data generated is not a straightforward process. Imposing specified priors such as anatomical, mathematical or functional constraints on the procedure are necessary in order to obtain a unique and probable answer to this problem.

In the absence of a structural scan, SPM's template head

model (based on the MNI brain) was transformed to match the fiducials in order to obtain a cortical mesh. During this step, data coregistration, each fiducial and each sensor position in the MEG space is matched to a corresponding one in MRI space. The next step, forward head model computation, computes the effect on the sensors of every dipole on the cortical mesh.

The individual M100 peaks identified in the previous chapter were used for dipole-modelling. After viewing the MMF scalp potential distribution, four dipoles were identified visually and defined a priori (see Figure 13). This observation matched the findings of previous studies using the mismatch paradigm, which used A1 and posterior STG of both brain hemispheres (Javitt, Steinschneider, Schroeder, & Arezzo, 1996; Opitz, Mecklinger, Friederici, & von Cramon, 1999; Schofield et al., 2009; Ulanovsky, Las, & Nelken, 2003).

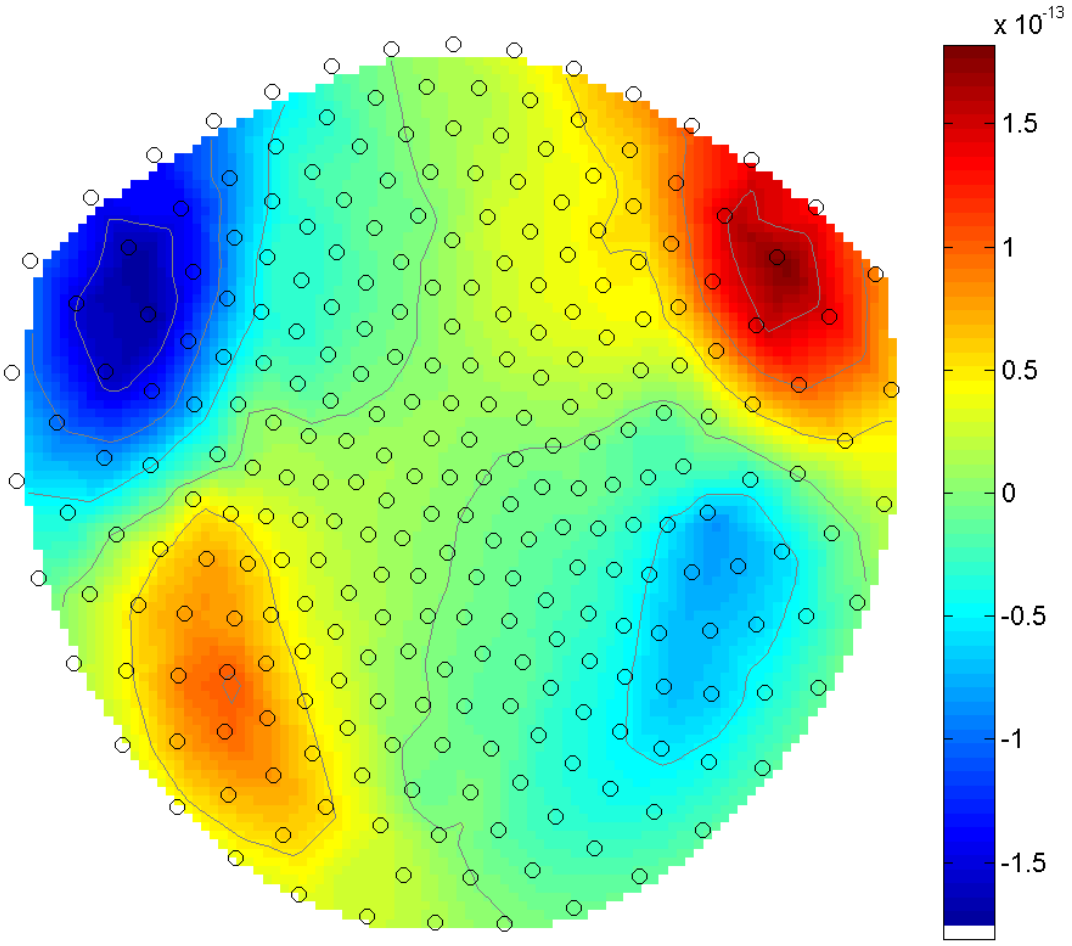

Figure 13: Example scalp plot for an individual subject (four dipoles)

In order to localise the equivalent current dipoles, an iterative Variational-Bayesian Equivalent Current Dipole/VB-ECD (Kiebel, Daunizeau, Phillips, & Friston, 2008) method was employed, which estimated the posterior distribution (P_{POST}) of the parameters, thus making the confidence interval of the estimated parameters directly available. With this method, models can be compared using their model evidence. Using VB-ECD, it is therefore possible to ask, for example, whether an MEG dataset could be modelled better with

three, four or five dipoles. Classic goodness-of-fit models are inferior to model comparisons as they do not take into account the complexity of models. This method fits equivalent current dipoles to a dataset, after specifying a mean and variance over source locations and moments.

VB-ECD is a Bayesian algorithm which requires a previous mean and variance over source locations and moments. The means for the locations of the dipoles were adopted from an fMRI study (Opitz, Rinne, Mecklinger, von Cramon, & Schröger, 2002) on mismatch negativity which used the following coordinates: right A1 (46, -14, 8), left A1 (-42, -22, 7), right STG (59, -25, 8), and left STG (-61, -32, 8). The previous variance in the position of the four dipoles was set to 100mm and used a prior variance of 100nA/m^2 on the dipole moments. The VB-ECD method was run for 100 different initialisations for each family of configurations (two, three and four dipole models). The model with the maximum model evidence for each family and for each subject was used for second-level analysis. In the L1 speakers, the average coordinates for left A1, right A1, left STG and right STG were (-43±24, -33±20, 1±21), (54±21, -16±27, -17±17), (-50±11, -32±13, -2±15) and (57±13, -32±22, -4±19), respectively. In the L2 speakers, the average coordinates for left A1, right A1, left STG and right STG were (-47±13, -27±18, -4±16), (48±20, -20±21, -4±14), (-60±9, -

36±11, 8±16) and (56±15, -29±17, 6±16), respectively (see Tables 5 and 6 for dipole coordinates for all individual subjects). Dipole positions for all subjects, L1 and L2 speakers separately were plotted onto an MNI glass brain (see Figures 14, 15 and 16).

subject	dipoles	X	Y	Z
1	Left A1	-54	-26	-1
	Right A1	64	-22	5
	Left STG	-29	-10	-2
	Right STG	66	-5	-9
2	Left A1	-37	-25	23
	Right A1	5	5	-41
	Left STG	-50	-46	11
	Right STG	68	-38	0
3	Left A1	9	-61	-40
	Right A1	54	0	-15
	Left STG	-46	-44	-5
	Right STG	58	-8	0
4	Left A1	-58	-45	13
	Right A1	70	-20	3
	Left STG	-59	-41	16
	Right STG	65	-12	9
5	Left A1	-62	-54	9
	Right A1	67	-10	-7
	Left STG	-39	-16	-8
	Right STG	61	-45	19
6	Left A1	-66	-23	-8
	Right A1	64	-16	-30
	Left STG	-57	-25	-33
	Right STG	62	-31	-13
7	Left A1	-60	-43	20
	Right A1	70	-26	-9
	Left STG	-51	-24	4
	Right STG	53	-56	20
8	Left A1	-24	2	14
	Right A1	48	-76	-20
	Left STG	-68	-42	-8
	Right STG	24	-70	-40
9	Left A1	-35	-19	-24
	Right A1	41	20	-39
	Left STG	-54	-39	11
	Right STG	54	-26	-20

Table 5: Dipole coordinates for L1 speakers

subject	dipoles	X	Y	Z
1	Left A1	-47	-45	-40
	Right A1	63	-23	-21
	Left STG	-59	-36	-13
	Right STG	41	-56	-4
2	Left A1	-59	-42	12
	Right A1	-7	-80	-8
	Left STG	-62	-38	-7
	Right STG	71	-29	7
3	Left A1	-25	-10	4
	Right A1	38	-11	-4
	Left STG	-67	-22	23
	Right STG	70	-22	16
4	Left A1	-67	-8	11
	Right A1	56	-15	14
	Left STG	-48	-16	27
	Right STG	38	-1	37
5	Left A1	-40	-18	6
	Right A1	69	-14	0
	Left STG	-63	-36	13
	Right STG	43	-60	14
6	Left A1	-24	0	14
	Right A1	38	-26	25
	Left STG	-61	-25	23
	Right STG	71	-31	2
7	Left A1	-47	-27	-13
	Right A1	54	-1	8
	Left STG	-62	-37	8
	Right STG	71	-26	-4
8	Left A1	-49	-32	-8
	Right A1	63	-17	-10
	Left STG	-66	-51	-14
	Right STG	56	-23	-25
9	Left A1	-43	-57	8
	Right A1	57	-29	-18
	Left STG	-43	-36	-20
	Right STG	30	-6	6
10	Left A1	-47	-36	-11
	Right A1	66	-13	-26
	Left STG	-44	-53	22
	Right STG	66	-35	6
11	Left A1	-64	-45	-27
	Right A1	48	1	-10
	Left STG	-66	-32	1
	Right STG	66	-35	3
12	Left A1	-53	-5	-4
	Right A1	42	-3	-2
	Left STG	-67	-39	22
	Right STG	64	-17	24
13	Left A1	-43	-29	-7
	Right A1	41	-22	5
	Left STG	-67	-48	14
	Right STG	42	-31	-10

Table 6: Dipole coordinates for L2 speakers

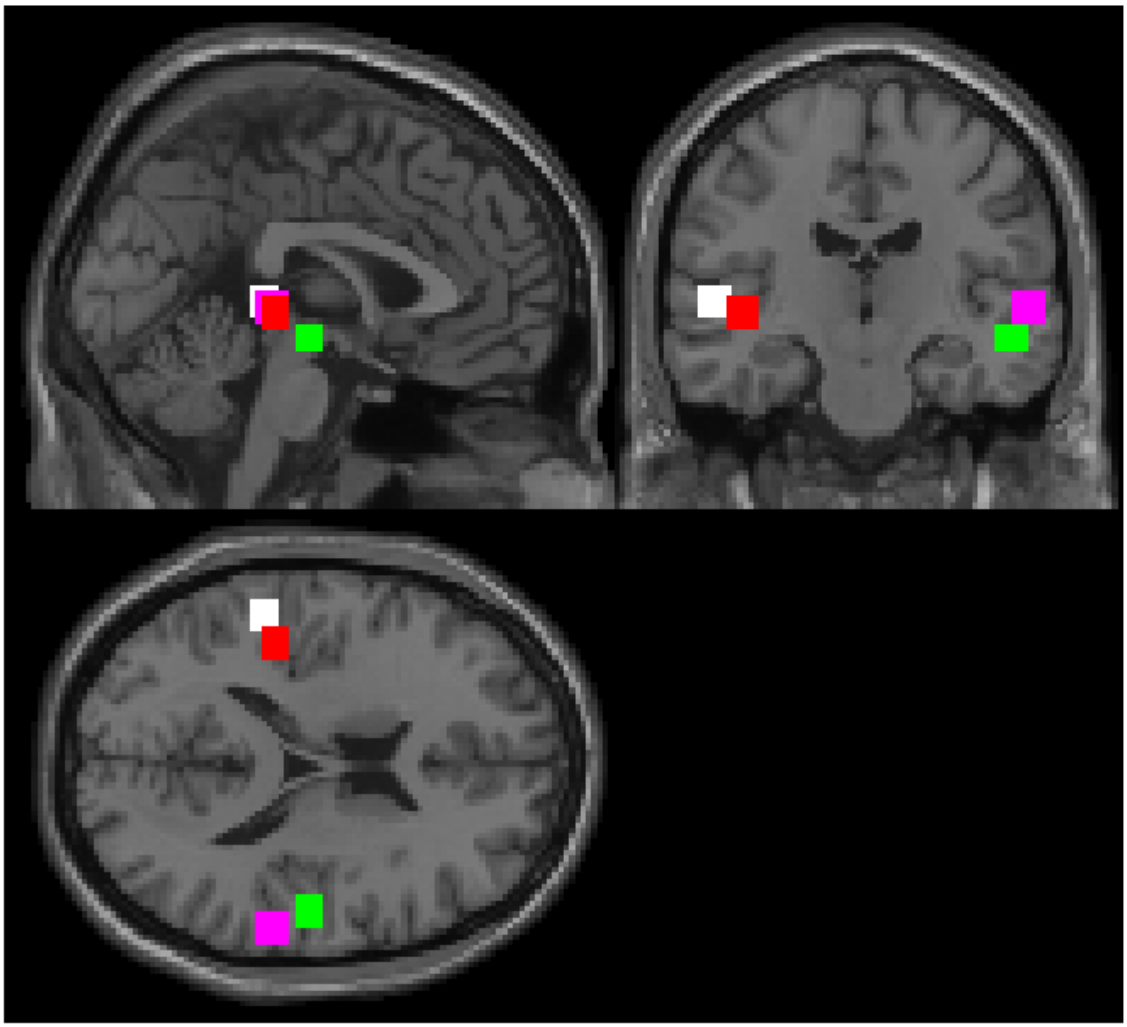

Figure 14: Dipoles for all of the subjects plotted onto an MNI glass brain (LA1=pink, RA1=green, LSTG=white, RSTG=red)

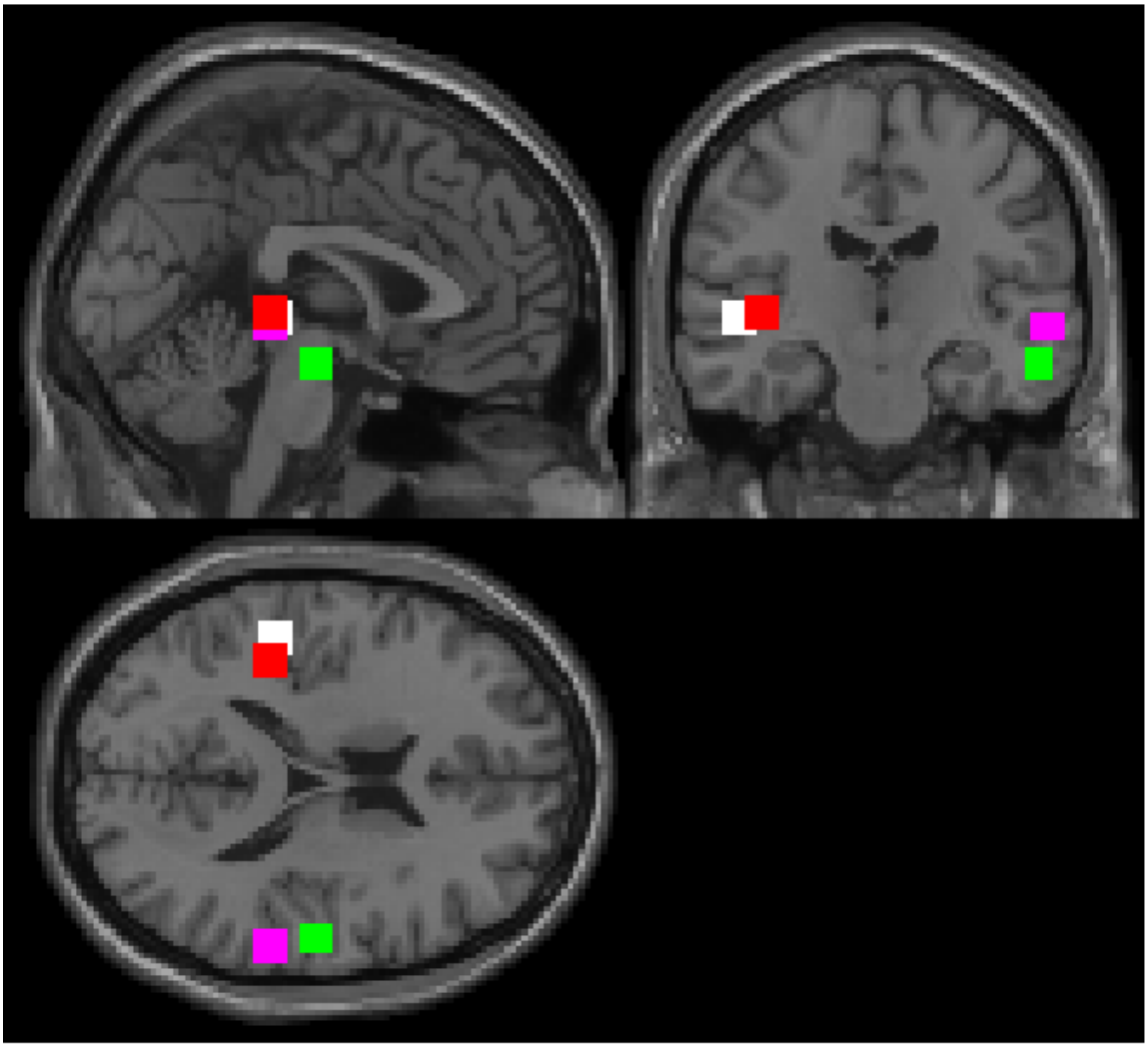

Figure 15: Dipoles for all of the L1 speakers plotted onto an MNI glass brain (LA1=pink, RA1=green, LSTG=white, RSTG=red)

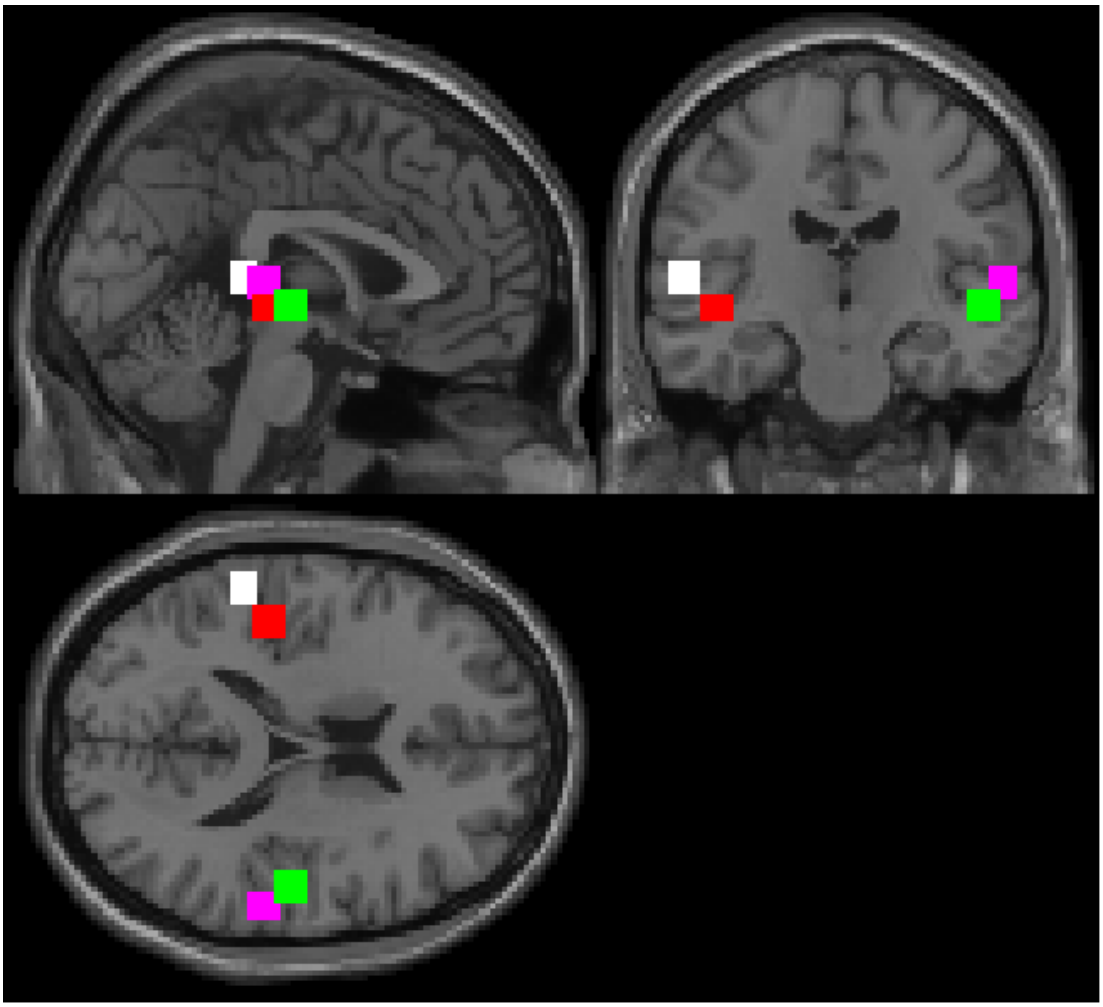

Figure 16: Dipoles for all of the L2 speakers plotted onto an MNI glass brain (LA1=pink, RA1=green, LSTG=white, RSTG=red)

Dynamic causal modelling

Dynamic causal modeling adds a neuronal model underneath the forward model. VB-ECD source reconstruction is used to localise active sources in the brain which are then used to specify location priors of sources in DCM. DCM modeling is based on the neural mass model (Felleman & Van Essen, 1991; Jansen & Rit, 1995) of the cortical column and posits three different types of connections in the brain: forward, backward and lateral connections. The neural mass model defines connections between layers of the cortex, according to known connectivity architecture. Forward connections, which form synapses in the granular layer of the cortex, have an excitatory effect. Backward connections, which form synapses in the supra- and infragranular levels of the cortex, excite both excitatory pyramidal cells and inhibitory interneurons. Lateral connections form synapses on all three layers of the cortex and can therefore have both excitatory and inhibitory effects.

Based on this connectivity pattern between the layers of the cortex and additional different neuronal dynamics for each layer, a generative spatiotemporal forward model (Kiebel, Garrido, Moran, Chen, & Friston, 2009; Kiebel, Garrido, Moran, & Friston, 2008) of the MEG data is specified. This model then estimates how well it fits the data. Additionally to the forward, backward and lateral connections mentioned above, DCM makes use of self-connections.

Self-connections (Kiebel et al., 2007) influence the amplitude of post-synaptic responses of cell populations.

Our study examines phonemic deviancy by exploring the connection strengths and the modulation of connections between four dipole sources for three deviants compared to the standard stimulus. Using a four dipole model, the number of all possible connections (16) specifies the number of different dynamic causal models (2^16-1=65535). This number of models is impractical due to computing constraints. This fact led us to constrain our model configuration by, firstly, dispensing with diagonal connections between auditory cortices and superior temporal gyri in opposing hemispheres and, secondly, adding fixed self-connections to each model. Adding to these connections two forward, two backward and four lateral connections (which could vary), we come to a sum of 255 models (2^8-1). All the models for each participant (255) were entered groupwise into a group-level Bayesian Model average analysis (Penny et al., 2010; Schofield et al., 2012; Stephan, Penny, Daunizeau, Moran, & Friston, 2009) with a random-effects design. This analysis calculated average modulations of connection strength for all connections and all models in the model space.

Not only the relative fit, but also the relative complexity of competing models must be considered when comparing different models (Pitt & Myung, 2002; Schofield et al., 2012). Comparisons

of competing models of different complexity can be offered by Bayesian model selection/BMS (Penny et al., 2004; Schofield et al., 2012; Stephan et al., 2009). We used a hierarchical method for BMS (Penny et al., 2010; Stephan et al., 2009; Teki et al., 2013), with a random-effects group design about the relative fit of different competing models of differing complexity for analysis. BMS uses variational Bayes to infer the posterior density of the models. The exceedance probability is then derived (i.e. the probability that a particular model is more likely than any other model). This analysis was used to calculate average connection strength for each connection across all models in model space. A non-parametric proportion test (Penny et al., 2010; Teki et al., 2013) was used to test statistical significance of the Bayesian Model Average results, connection by connection.

Results

In our study the four dipole model had the best model evidence when compared with the two or three source models. Estimated frequencies with which these models are used in the population are represented by expected posterior probabilities (P_{POST}). As expected, following previous MEG and fMRI studies (Schofield et al., 2009; Schofield et al., 2012; Teki et al., 2013), the four dipole model was the winning model family for the mismatch data for both

groups (see Figure 17). Our results (all deviants) were all best explained by a four dipole model. This result was not just due to the fact that the four dipole model was the most complex model tested: Bayesian model comparison takes the complexity of a model into account when explaining the data and does not overfit the data (Schofield et al., 2012).

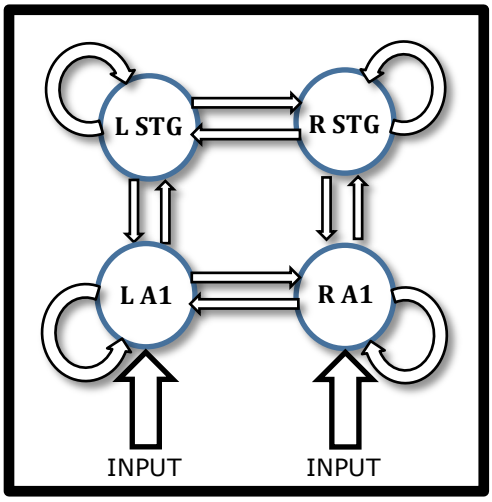

Figure 17: Winning four dipole source model with forward, backward and lateral connections

As the four dipole model was the model which best explained the MEG responses to standards, we then used these four dipole sources to evaluate the responses to both standards and deviants over the entire time period (0–300 ms).

We performed a multivariate dynamic causal modelling analysis in which interactions between regions were expressed over the whole of peristimulus time (0–300 ms). The connection strengths of the evoked difference waves for the three deviants

within a window of 150–250 ms were measured for each dipole within the winning model and we then performed a mixed model analysis, using the connection strengths for each individual subject for D1, D2 and D3.

1) All subjects analysis of D1, D2 and D3

In the first step, we observed the Bayesian Model average results for all subjects, divided into the deviants 1, 2 and 3 (see Figure 18). Significantly modulated connections are marked as red. A connection is considered relevant if the connection strength/c.s.≥±0.1. When the exceedance probability (P_{POST}) is >0.9 or <0.1, a connection is then considered significantly modulated. As we can see, compared to the standard, the within-category deviant D1 has two significantly modulated connections compared to the standard stimulus: The left and right hemisphere STG self-connections are activated more (c.s.=1.13, P_{POST}>0.9 and c.s.=1.14, P_{POST}>0.9 respectively) when processing the mismatch. D2 has four significantly modulated connections compared to the standard stimulus: The LA1 self-connection (c.s.= 1.14, P_{POST}>0.9), the RA1 self-connection (c.s.=1.18, P_{POST}>0.9), the LSTG self-connection (c.s.=1.16, P_{POST}>0.9) and the lateral LA1 to RA1 connection (c.s.=1.16, P_{POST}>0.9) are activated more. D3 has four significantly modulated connections compared to the standard stimulus: The LA1 and RA1 self-connections (c.s.=1.15, P_{POST}>0.9

and c.s.=1.11, P_{POST}>0.9 respectively), the RSTG to LSTG lateral connection (c.s.=1.29, P_{POST}>0.9) and the LA1 to RA1 lateral connection (c.s.=1.16, P_{POST}>0.9) are activated more.

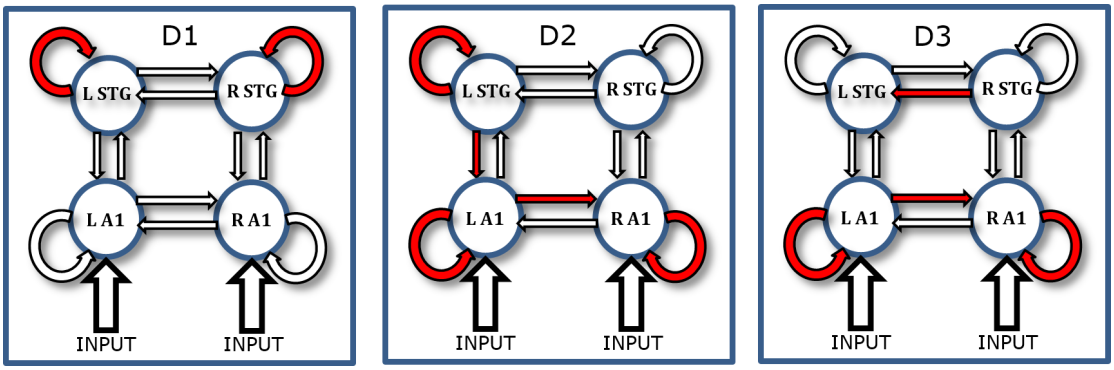

Figure 18: All subjects analysis for D1, D2 and D3

When looking at D2, the between-category deviant for L1 speakers and within-/between-category deviant for L2 speakers, we can see that several more connections are significantly modulated compared to the standard. There is additional left (c.s.=1.14, P_{POST}>0.9) and right (c.s.=1.18, P_{POST}>0.9) auditory cortex self-connection and left superior temporal gyrus self-connection involvement (c.s.=1.16, P_{POST}>0.9). Addtionally, there is a top-down/feedforward connection between left STG and left A1 (c.s.= 1.17, P_{POST}>0.9) and a lateral connection between left and right A1 (c.s.=1.16, P_{POST}>0.9).

When looking at D3, the between-category deviant for both L1 and L2 speakers, compared to the standard there are significantly modulated self-connections in left and right A1 (c.s.= 1.15, P_{POST}>0.9 and c.s.=1.11, P_{POST}>0.9, respectively) and additional lateral connections: right to left STG (c.s.=1.29,

P_{POST}>0.9) and left to right A1 (c.s.= 1.16, P_{POST}>0.09). So, in sum, the more deviant a speech sound is perceived to be compared to the standard, the more the involvement of self-connections and the more connections between brain regions.

To look at this effect in more detail there follow simple effects comparisons of the deviants (see Figure 19). In the D3-D1 comparison, both the LA1 (c.s.=0.10, P_{POST}>0.09) and RA1 (c.s.=0.12, P_{POST}>0.09) self-connections are significantly modulated. What this means is that increasing deviance is coupled with an increasing strength of significantly modulated self-connections in the auditory cortex.

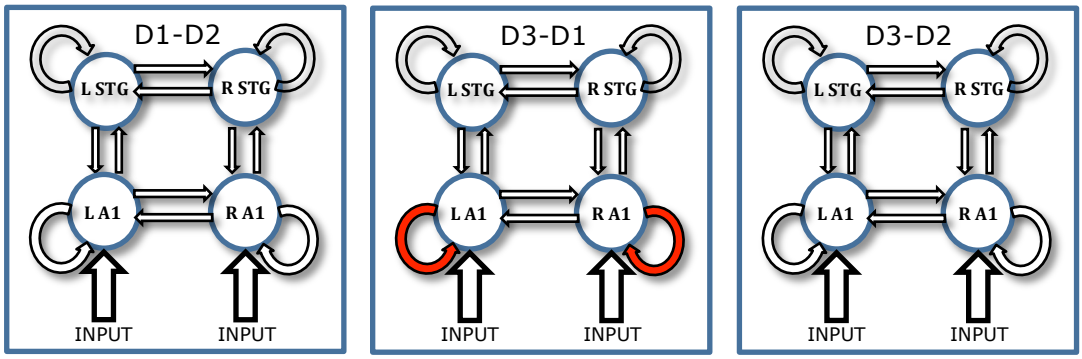

Figure 19: Simple effects comparisons of the deviants for all subjects

2) English-French comparisons

Below there are similar figures as above, however here we are subtracting French subject scores from English subject scores. As can be seen (see Figure 20), there is a significant effect of language: L2 speakers show greater self connections for right STG

for D2 (c.s.=-0.26, P_{POST}<0.1) and D3 (c.s.=-0.22, P_{POST}<0.1). There are no significant simple effects, however there does seem to be a generally greater use of right STG for L2 speakers, regardless of deviance.

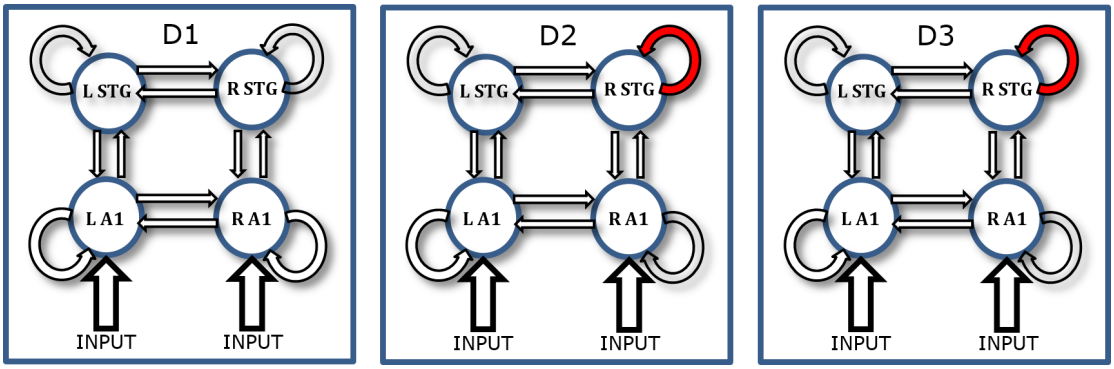

Figure 20: French-English interactions

3) Comparisons with behavioural data

In our previous conventional sensor-space analysis there were no significant English-French main effects, however, there were significant correlates with perceptual behavioural measures (in particular, category discrimination for beat-bit). Due to the varying L2 abilities in the L2 speaker group, the behavioural measures may be more sensitive. Using a mixed model analysis, we investigated the mean connection strength for each subject per deviant (for connections which were significant in the overall analysis) and the STG self-connection in cases with cross-language differences. Behavioural measures used were category discrimination for beat-bit, average phoneme identification for the equated-duration stimuli, acoustic separation of produced beat and bit (production)

and acoustic discrimination of beat and a close pair. The analysis was performed across all subjects (L1 speakers and L2 speakers mixed). Three connection strength values were discarded due to being extreme outliers.

D1

There were no significant behavioural correlates with D1. Both L1 and L2 speakers heard this deviant as a within-category deviant.

D2

There were differences between the two groups for the deviant D2 and there were also significant behavioural correlates: category discrimination (beat-bit) is related to the right STG self-connection ($F(1,17)=7.32$, $p<0.05$), in that the better the category discrimination abilities, the weaker this connection. ID relates to the left A1 to right A1 connection ($F(1,17)=9.25$, $p<0.05$), in that the better the ID abilities, the weaker the connection. ID additionally is related to the left STG to left A1 connection ($F(1,17)=7.06$, $p<0.5$), in that the better the performance on ID, the stronger the connection. In sum, for D2, a better performance on beat-bit is related to weaker lateral connections towards the right hemisphere or within the right hemisphere, and stronger connections within the left hemisphere.

D3

As was the case for D1, D3 was categorised the same by the two groups of speakers. Everyone could identify the beat-boot distinction without problems and there were no significant behavioural correlates with D3.

4) Comparisons to population measures

Using a mixed model analysis, we investigated the mean connection strength for each subject per deviant (for connections which were significant in the overall analysis) and the population measures (for L2 speakers only) which were collected using a language background questionnaire. Population measures were: age, age of acquisition and length of exposure to English in the UK/residency in the UK.

D1

Age correlated with D1 and the left A1 self-connection ($F(1,5)=40.37$, $p<0.05$) and background correlated with significant D1, D2, D3 overall connections.

D2

For the L2 speakers, their age of L2 acquisition correlated with D2 and the right STG self-connection ($F(1,5)=8.27$, $p<0.05$), as did their UK residency/exposure to English in the UK ($F(1,5)=6.66$, $p<0.05$).

D3

Age and residency correlated with D3 and the right STG self-connection ($F(1,5)=9.75$, $p<0.05$).

Discussion

We were interested in finding out whether L2 speakers show differences in their neural architecture when listening to L2 speech sounds, as compared to L1 speakers. Assuming that native and non-native speakers hear the same speech sounds differently initially and that some non-native speakers can perform to native-like standards (Birdsong, 1999) in specific linguistic domains, there must be some point in time or a certain amount of L2 exposure which produces a difference in how non-native speakers hear or categorise non-native speech sounds.

At some point, behaviourally, these L2 speakers can become highly proficient. We would like to know what this auditory shift looks like in terms of brain architecture – is it visible in terms of changed connectivity between brain regions? Are there differences in brain architecture between L1 and L2 speakers? The standard univariate mismatch negativity analysis in the previous chapter did not indicate significant differences between L1 and L2 speakers. The subsequent dynamic causal modeling approach on the same data, however, did show significant differences between the two groups.

This is probably due to the fact that the classic sensor-space approach does not take into account the entire mismatch difference wave generated, but only concentrates on the mismatch peak picked by the researcher, discarding the rest of the data. DCM, however, models using all of the data generated, which also includes differences in neural processing between L1 and L2 speakers which are expressed earlier than in the classic mismatch peak.

In fact, comparatively early modulation due to deviancy is already observed within the first 50ms after sound onset (Grimm & Escera, 2012). This early modulation is due to stimulus-specific adaptation, whereby neurons reduce their responsiveness to commonly heard sounds while they remain sensitive to rarely heard sounds. This phenomenon enhances the saliency of surprising sounds against a background of standard sounds (Ulanovsky et al., 2003). Neurons which show stimulus specific adaptation exist at severel levels of the auditory pathway, from the inferior colliculus up to the auditory cortex (Aguilar Ayala & Malmierca, 2013). This means that DCM (which analyses the whole response) will catch this early modulation while peak picking will not.

Additionally, DCM goes beyond sensor-space analysis and focuses on source-space analysis. Through inversion of the forward model, source-space analysis and Bayesian model estimation, DCM

provides additional information as to how many dipoles are involved in neural processing, how these dipoles are connected, and how these connections change, depending on deviant type and language background of the listener. We know that all speakers (L1 and L2) show a classic sensor-space mismatch response for each deviant. What DCM adds to these expected findings is a way of seeing what underlying neural speech sound processing is happening as a cause for these externally visible mismatch responses.

However, there are limitations to the interpretation of DCM results in general, especially when discussing connectivity effects. Interpreting connectivity effects can be challenging due to DCM generating highly complex data patterns. Interpretation of complex DCM interactions may therefore be speculative in nature. Especially when looking at changes in connections and their effects on behaviour it is not always possible to say which change specifically has an effect or whether it is a combination of changes which leads to an effect.

Our winning model had four dipoles (as in Schofield et al., 2009; Teki et al., 2013). This was the maximum amount of possible dipoles we had allowed for. A similar study by Schofield et al. (Schofield et al., 2009) on the MMF, using very similar deviants and experimental design, had concluded that a five dipole model (with additional right/left IFG) did not fit the MMF data well, despite being

a more complex model.

When looking at the data for all subjects (L1 and L2 speakers together), D1, the within-category/acoustic deviant, elicited additional left and right hemisphere connections and additional STG self-connections. According to predictive coding theory, the STG self-connections point to higher-level processing of the acoustic/within-category MMN after an error signal generated by mismatch between expected and actual input to A1. Self-connections might also mean that the error signal generated is large. D2, the within-/between-category deviant elicited several more connections compared to the standard: there was additional left and right A1 self-connection involvement, top-down/feedforward connection between left STG and left A1 and a lateral connection between both A1s. The STG involvement points to changes in the higher levels of the neuronal network due to the violation of phoneme expectation (built up over years of experience). This MMF data includes both L1 and L2 speakers, so we need to explore the difference in connections between L1 and L2 speakers. D3, the between-category deviant, elicited self-connections in left and right A1 and additional lateral connections: right to left STG and left to right A1. The more deviant a speech sound is perceived to be, compared to the standard, the more the involvement of self-connections and the more connections between

brain regions. Increasing deviance is coupled with an increasing strength of significant self-connections in A1. According to Schofield et al. (Schofield et al., 2009), there is a left hemisphere bias for phoneme deviants, however, we do not seem to see this bias in our DCM data (although a left hemisphere bias effect is visible in our MMN peak data in Chapter 2).

After looking at the data for all subjects together, we investigated the L1-L2 speaker interactions. We saw an effect of language, in that French L1 speakers show greater self connections for right STG for D2 and D3. The existence of right STG self-connections for D2 and D3 (as compared to the L1 speakers) seem to indicate that the L2 speakers are having more problems categorising L2 phonemes than the L1 speakers. The right hemisphere seems to be helping out the left hemisphere which can not cope with processing. There seems to be a generally greater use of right STG for French L1 speakers, regardless of deviance.

According to several studies (Luo et al., 2006; Naatanen et al., 2007; Teamu Rinne et al., 1999; Tervaniemi & Hugdahl, 2003; Zatorre, Belin, & Penhune, 2002), the left hemisphere is more specialised in processing temporal information and the right hemisphere in processing spectral information. All of our deviants differed from the standard in the frequency dimension, so our data could tie in with this theory, in that the error signal produced for the

L2 speakers could be being processed on the basis of frequency. However, a difference between L1 and L2 speakers in the involvement of the left hemisphere and the right hemisphere is actually evidence against this theory.

Indeed, there is alternative way of interpreting our results. According to Hickok and Poeppel and their dual stream model (Hickok, 2013; Hickok, 2012; Hickok & Poeppel, 2007), based on lesion data, both hemispheres work in parallel on different time-scales by two different processing streams. Their theory states that the right hemisphere is specialised for longer time-scales (slower rate sample processing) such as suprasegmental stimuli (ie. syllable-boundary and syllabic-rate cues, lexical tonal information, prosodic and stress cues) and faster rate sample processing for shorter time-scales such as segmental stimuli (ie. phonemes) is bilaterally shared.

When adding DCM to this discussion of hemisphere specialisation, dominancy, and biases, one could see DCM as building a bridge between these two different theories. These two theories investigate tradtional functional brain imaging of auditory stimuli without computing neuronal sources and functional connectivity between brain regions. Looking at our MEG data, one can see that it is entirely possible for the same MEG data to show hemisphere biases in one type of analysis (traditional) and then

show bilateral changing connectivity between brain regions depending on native-speaker status in the other (using DCM). This leads us to the conclusion that speech perception may be bilaterally organised, but with DCM computed differences in bias which depend on the type of stimulus and the subject.

When looking at the correlations with the behavioural data, conventional sensor-space analysis showed no significant English-French main effects. However, with DCM source space analysis, we see significant correlates with perceptual behavioural measures (in particular, category discrimination for beat-bit). D1 showed no significant behavioural correlates. Both L1 and L2 speakers heard this deviant as within-category deviant. The same is true for D3, which was categorised the same by both two groups of speakers: everyone could identify the beat-boot distinction without problems. However, when looking at D2 and its relationship to behavioural measures, the within-/between-category deviant (depending on language ability), there were differences between speakers, depending on their ability to categorise English speech sounds.

When looking at the relationships between the behavioural measures and the DCM D2 mismatch response, we see that speakers show some significant behavioural correlates to D2, in that category discrimination for beat-bit is significantly related to the right STG self-connection. Speakers who are low ability speakers

show more right STG self-connection activity, indicating their problems in categorising L2 phonemes. These low ability L2 speakers have not yet developed a good phoneme category for "bit" in their L2.

According to predictive coding theory, an error signal is produced when these low ability L2 speakers hear the L2 phoneme ("bit"). This error signal is based on some kind of higher-level neural processing which tries to integrate and adapt the sound heard into the existing incomplete L2 phoneme inventory. A study by Schofield et al. (Schofield et al., 2009), with a very similar paradigm, stimuli and DCM analysis, looked at how speech and non-speech stimuli were processed in a neural network. Their English L1-speaking subjects' brains differentiated between speech and non-speech (perceptually matched tone) stimuli, in that speech stimuli elicited more activation in the left hemisphere (LSTG) and non-speech stimuli elicited more activation in the right hemisphere (RA1). Following from this, this could suggest that our D2 stimuli were being processed in a more non-speech manner than D1 and D3 for the L2 speakers.

We can see a similar setup for phoneme identification abilities. Low ability speakers show a stronger left A1 to right A1 lateral connection. The better the identification abilities, the weaker this connection. This seems to show that low ability L2 speakers

rely more on right-hemisphere components for processing speech mismatches. These results are comparable to Teki et al.'s (Teki et al., 2013) findings: in their study with aphasics and non-aphasics, with a very similar paradigm, stimuli and DCM analysis, in that phoneme deviants in Teki et al.'s study also elicited left A1 to right A1 activity for aphasics. In this aspect, aphasics and L2 speakers seem to have comparable activation. Aphasics could be described as lacking the normal speech perception abilities of a native speaker (through injury to the brain). Low ability L2 speakers also lack the normal speech perception abilities for the language they are listening to. The right hemisphere seems to be compensating for or aiding the left hemisphere.

Additionally, in our study identification abilities are related to the left STG to left A1 connection: the better the performance on identification, the stronger this connection. This could indicate that low ability L2 speakers have not yet managed to integrate the L2 speech sound into their new L2 phoneme inventory, as, according to predictive coding theory, increased activation of the LSTG is supposed to be linked to adapting predictions and expectations to actual sounds heard. They might be processing it more with their right hemisphere, as a non-speech sound or frequency based. Over time and exposure to the new L2 sound, we would expect this connection to be strengthened with frequent use and to come to

resemble the strong and efficient LSTG to LA1 connection shown by high ability L2 speakers and L1 speakers. In sum, a better performance on beat-bit categorisation and identification is linked to weaker lateral connections towards the right hemisphere or within the right hemisphere and stronger connections within the left hemisphere.

We looked at the population measures (for the L2 speakers only), however this did not add any additional knowledge – the L2 speakers' age of acquisition of English as an L2 correlated with their right STG self connection. However, as this is indirectly also a measure of their length of exposure to English and UK residency (roughly, exposure to L2 in a naturalistic environment) and exposure also correlates with the right STG self connection, this is most likely just another way of measuring their L2 identification and categorisation abilities, in that individuals who are worse with identifying and categorising beat-bit have stronger right STG self connections.

Summary

DCM modeling showed that a four dipole model was the best fit for the MEG mismatch data. This ties in with previous research on phonemes (Schofield et al., 2009; Teki et al., 2013). The mismatch for each deviant showed a differing neural architecture, compared

to the standard. Increasing deviance leads to increasing involvement of self connections and lateral connections.

As expected, both L1 and L2 speakers performed similarly for the within-category/acoustic deviant 1. However, when it came to D2 and D3, the L2 speakers had greater self connections for RSTG. Normal processing strategies for the L2 speakers are not adequate when confronted with a not fully represented L2 speech sound, therefore additional processing by the right hemisphere must take place.

The behavioural measures linked to D2 only – D1 and D3 were processed similarly by L1 and L2 speakers. Low ability L2 speakers activated the RSTG self connection more strongly during category discrimination for D2. Low ability L2 speakers activated the left A1 to right A1 lateral connection more, suggesting that they are relying more on the right hemisphere to process this mismatch response. In this aspect, they are functioning in a similar manner to aphasics (Teki et al., 2013). Additionally, low ability L2 speakers activated the top-down LSTG to LA1 connection less than high ability L2 speakers during identification. This could indicate their less efficient processing procedure when faced with a not yet known/stabilised L2 sound and reliance on right hemisphere based processing instead. The population measure of residency in the UK also indicated that low ability L2 speakers have stronger right STG

self connections, showing that time and exposure have a favourable influence on neural processing. The fact that high ability L2 speakers are indistinguishable from L1 speakers at the neural architecture level for D2 indicates that the human brain is highly adaptable and retains plasticity in adults by learning to identify and categorise a novel sound and by sucessfully integrating it into an automatically functioning phoneme inventory.

Chapter 4: The Acoustic Change Complex

Introduction

After using the Magnetic Mismatch Response and Dynamic Causal Modeling to investigate L2 vowel processing we now turn to the Acoustic Change Complex as an experimental paradigm for exploring L2 vowel processing. The Acoustic Change Complex is a set of components which are elicited through sound change. It is a P1-N1-P2 complex (from now on referred to as the Acoustic Change Complex or ACC) which is related to another component complex, the onset P1-N1-P2. The onset P1-N1-P2 was first discovered in 1939 (Davis, 1939). It is elicited by the presence of sound after silence and is primarily clinically used to determine whether sounds are detected by the auditory cortex, for example for hearing tests. The ACC is a cortical auditory evoked potential (Steinschneider, Liegeois-Chauvel, & Brugge, 2011), as is the MMN. It is elicited by sound changes and is thought to represent neural detection and discrimination of the acoustic properties of sounds; the ACC is elicited by frequency and intensity changes in synthetic vowels (Martin, Tremblay, & Stapells, 2007) and spectral and intensity changes in speech-like sounds (Ostroff, Martin, & Boothroyd, 1998).

The ACC is different to the onset P1-N1-P2 response, in that the ACC only occurs due to changes in continuous stimulus trains instead of after an initial sound onset. The first component, P1, is a

positive-going wave at around 50ms, the second component, N1, is a negative-going wave at around 100ms and the third component, P2, is a positive-going wave at around 200ms (Martin, Tremblay, & Korczak, 2008), see Figure 21.

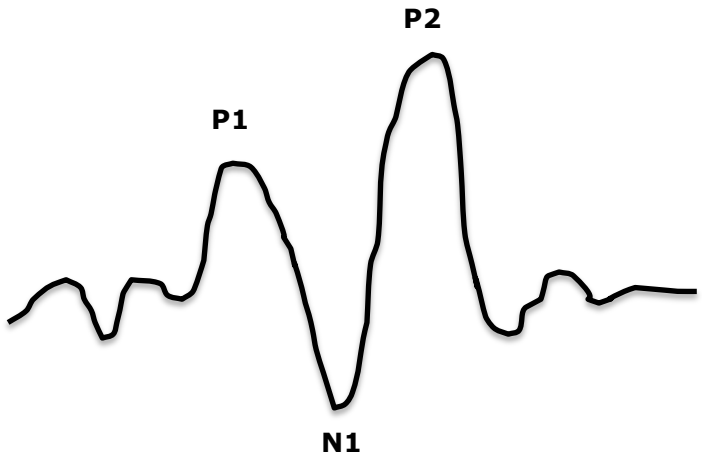

Figure 21: Typical ACC response with the P1, N1 and P2 components. The positive values are plotted upwards.

It is not clear whether the onset P1-N1-P2 and the ACC share the same generators and whether they represent the same neural processes (Martin et al., 2008). However, there seems to be a consensus as to the location of the P1-N1-P2 generators in the brain.

The P1 is an obligatory sensory response upon the detection of sound input entering the system. The P1 neural generators are in the primary auditory cortex (specifically, Heschl's gyrus), the hippocampus, the planum temporale, the lateral temporal regions and possibly the subcortical regions (Liégeois-Chauvel, Musolino,

Badier, Marquis, & Chauvel, 1994). In adults the P1 often has only a small amplitude and N1 and P2 dominate the response (Martin et al., 2008).

The N1's neural generators are located in the primary and secondary auditory cortex (Martin et al., 2008; Scherg & von Cramon, 1986). The N1 consists of three underlying components: Firstly, the frontocentral negativity component N1b (Näätänen & Picton, 1987), is generated by bilateral vertically oriented dipoles in or near the auditory cortex in the superior temporal plane. The elicited response is largest near the Cz electrode (Vaughan Jr & Ritter, 1970) and is believed to be related to attention (Näätänen & Picton, 1987; Picton, Hillyard, Krausz, & Galambos, 1974). Secondly, the T-complex component (Näätänen & Picton, 1987; Tonnquist-Uhlen, Ponton, Eggermont, Kwong, & Don, 2003) is a negative wave at 70-80 ms (referred to as the Na), a positive wave at ca. 100ms (the Ta), followed by a larger negative wave at ca. 140-160ms (the Tb). It is generated by a radially oriented generator in the secondary auditory cortex, in the superior temporal gyrus. The T-complex is largest at midtemporal electrodes (Tonnquist-Uhlen et al., 2003), dipole source modeling has fitted one source to this component. Thirdly, the negative wave at 100ms (Näätänen & Picton, 1987), the N1. The N1 is largest at the Cz electrode. The N1 is different to the N1b component in that it is

most sensitive to long (more than 4s) interstimulus intervals. The N1 is affected by stimulus rate in that N100 amplitudes tend to be smaller for rapid stimulus rates (Stapells, 2002).

The P2's neural generators are located in multiple areas; Heschl's gyrus, primary auditory cortex and the secondary cortex (Crowley & Colrain, 2004; Hari et al., 1987; Scherg & von Cramon, 1986). Dipole source modeling has not managed to fit a single source to the P2. It is therefore thought that several distributed generators may contribute to the P2 response and MEG data suggests that the generators for the P2 response lie in the planum temporale and in Brodmann area 22 (the auditory association complex) (Crowley & Colrain, 2004). It is possible that multiple sources, centred around Heschl's gyrus, are responsible for P2m (Crowley & Colrain, 2004).

The ACC is indicative of the beginning of cortical sound processing (Hillyard & Kutas, 1983). It is an obligatory cortical response and reacts to changes in an ongoing sound or to a change in an auditory environment. It is known to be an objective index of an individual's auditory discrimination abilities and is sensitive to small changes in F2 (Martin & Boothroyd, 2000). The ACC detects changes in frequency, amplitude and periodicity in consonant-vowel syllables (Ostroff et al., 1998), changes in amplitude and formant frequency within vowels (Martin & Boothroyd, 2000) and it shows

good agreement with behavioural discrimination performance (Martin et al., 2007). Complexity and duration of the stimulus can increase the N1 latency (Martin et al., 2007).

Almost all the research so far done on the ACC has been in the area of L1 auditory processing. To the best of my knowledge, there have been only two other studies so far in the field of L2 phoneme processing using the ACC. Brint (Brint, 2012) investigated /r-l/ consonant processing with English L1 and Japanese L1/English L2 speakers, but found no cross-language differences. Martin et al. (Martin, Shafer, Wroblewski, & An, 2012) investigated one vowel pair with English L1 speakers and Spanish L1/English L2 speakers. They found cross-language differences for the topography of the P2 component.

However, there have been some studies with L2 speakers investigating the onset P1-N1-P2 and the T-complex. Recent research (Wagner et al., 2013) measured the onset P1-N1-P2 and the T-complex with English L1 speakers and Polish L1/English L2 speakers who had started with English as an L2 after the age of 15, looking at the effect of language background. Their stimuli were nonsense word pairs which featured a Polish phoneme cluster onset (which is not a permitted English phoneme word onset cluster) or an English phoneme cluster onset (which is a permitted phoneme word onset cluster for both English and Polish). Their research

indicated that the onset P1-N1-P2 and the T-complex are both sensitive to phonemes, even though the L1 speakers could not discriminate behaviourally between the very similar Polish onset clusters and English onset clusters (/pt/ versus /pət/). However, only the T-complex reflected the difference in language background of the subjects.

Another study (Elangovan & Stuart, 2011), which looked at the onset P1-N1-P2 elicited in response to varying voice onset times with English L1 and Spanish L2 speakers, also did not show an effect of language background. Thus, changes in the morphology of the onset P1-N1-P2 response seem to be sensitive only to acoustic changes in speech sound, do not seem to be related to language background and seem to be independent of speech sound categorisation within languages.

The ACC paradigm is very efficient as an experimental paradigm because we can have a large amount of trials and elicit many more responses in a much shorter time than when using the mismatch response paradigm. A standard mismatch response oddball paradigm has a ratio of 4:1 of standards to deviants. At least two standards must be presented before a deviant may occur in order to elicit a mismatch response. However, a standard ACC paradigm can elicit an ACC response for every single sound change. ACC is therefore a very useful paradigm for taking a much broader

look at the perceptual vowel space than has been tried so far as many more trials can be fit into the same amount of time necessary for a mismatch response paradigm.

Apart from offering us the opportunity to collect large amounts of data in a relatively short time, the ACC also gives us an earlier view of auditory processing in speech. The ACC occurs at a much earlier point in time post-stimulus onset than the mismatch response. This is ideal for our purposes, as we plan to investigate whether language background has an influence at an even lower automatic auditory processing level than the MMN, however, at a higher level than the Auditory Brainstem Response.

Additionally, the design of our EEG studies is geared towards attempting the geometric mapping of L1 and L2 vowel perception with multidimensional scaling, as used in previous studies by Iverson et al. and Kuhl et al. (Iverson, Mulyak, & Wagner, 2013; Iverson et al., 2011; Kuhl & Iverson, 1995). We aim to map all vowels for the L1 and L2 speakers by using vowel stimuli which cover the whole vowel space for both languages. We thus hope to achieve two goals with our EEG experiments: Collecting ACC data and performing multidimensional scaling for all vowels for the L1 and L2 speakers.

We decided to use English native controls and German L1/English L2 speakers as subjects for these experiments. There are

some English production and perception difficulties (for example a/e confusion problems) which are typical for many German L1 speakers. This led us to believe we could possibly find differences between L1 and L2 speakers in brain responses as well as their behavioural measures.

We recruited 15 English native controls and 16 German L1/English L2 speakers of varying proficiency. It would have been interesting to have used French L1/English L2 speakers for this experiment too so that the MMN, DCM and ACC results would be comparable. However, this was not possible for practical reasons. The MMN and ACC experiments featured in this thesis were conducted 3-4 years apart and the original French native speakers were not available for participation anymore due to the majority of them having moved back to France in the meantime. Recruiting a large enough number of French native speakers for the MMN experiment had proven to be very time-consuming and difficult. This is why I decided to recruit German native speakers for this experiment instead as I knew many Germans in London and could use this network to recruit suitable subjects.

Our investigation into the Acoustic Change Complex started with an experiment collecting behavioural measures on L1 and L2 vowel perception from all L1 and L2 speakers who participated in the ACC experiments. This was in order to obtain accurate

behavioural data which could later be analysed, compared to and linked with the ACC responses. The stimuli were Southern British English and Standard German (=Hochdeutsch) vowels which covered the whole vowel space for both languages.

As with the MEG experiment, we acquired first the EEG data and then the behavioural data from each subject in order to not alert them to the purpose of the experiment before acquiring EEG data. All subjects in all experiments were tested individually and provided signed consent. They were administered a questionnaire assessing language, language history and basic information. All subjects were right-handed and reported no hearing or visual impairment (when wearing visual aids such as contact lenses or glasses), language impairment, learning or neurological disorders and were aged between 21 and 41 years old.

Experiment 1

The aim of this first experiment was to provide accurate behavioural L1 and L2 vowel processing data necessary for aiding interpretation of the ACC responses from the following two EEG experiments. This type of behavioural experiment has been proven to be an accurate measure of cross-language differences (Iverson et al., 2003) and is therefore useful for interpreting the L2 ACC responses. Subjects performed a combined vowel assimilation and vowel goodness

rating task.

Method

Subjects

The subjects were fifteen Southern British English L1 speakers, six females and nine males, who were between 22 and 41 years old (mean 28.6 years old, SD=5.6), and sixteen German L1 speakers, nine females and six males, who were between 21 and 33 years old (mean 26 years old, SD=3.4). The L2 speakers consisted of eleven subjects from North Germany and four subjects from South Germany. The German L1/English L2 speakers had had between 5 and 66 months of exposure to British English in the UK, living in London or in Southern England (mean 23.6 months residency, SD 19.3).

I had originally intended to test an equal amount of Northern German and Southern German speakers to see whether regional accent had an effect on results in behavioural and brain measures. However, recruiting and testing enough Southern German speakers who met my requirements (right-handed, no hearing/neurological impairments and brought up monolingually) would have taken more time than available. Having only four Southern German speakers made statistical analysis as a separate group impossible due to the small group size.

Stimuli

Subjects performed a combined assimilation and goodness rating task. The stimuli were 11 Southern British English and 11 North German vowels. The English and German stimuli were recorded in a word context by a bilingual Southern British English/North German female speaker in a sound-proof booth and were then further processed. The English vowel stimuli were: Boot, bead, bit, book, bed, bought, bad, bod, bird, bud and bard. The German vowel stimuli were: Spaet, huete, rad, schule, bitte, lied, kapUtt, flott, boese, steht and fett (see Figure 22 for all stimuli). The stimuli were naturally produced with the vowels of those two languages. We then cut the vowels from the words by first choosing a representative part of each vowel and then selecting two cycles centred around a vocal fold closure (from closure to closure at the zero-crossings, using PRAAT software).

The vowel stimuli had to be static as we wanted to elicit an ACC response to changes between vowel stimuli and not to changes within a vowel stimulus (due to fluctuation of F0). Therefore, we then concatenated these cycles to produce a longer, static vowel. In order to make the processed vowels sound more natural, the amplitude, duration and pitch contour of a natural production by the same speaker was imposed upon the vowels.

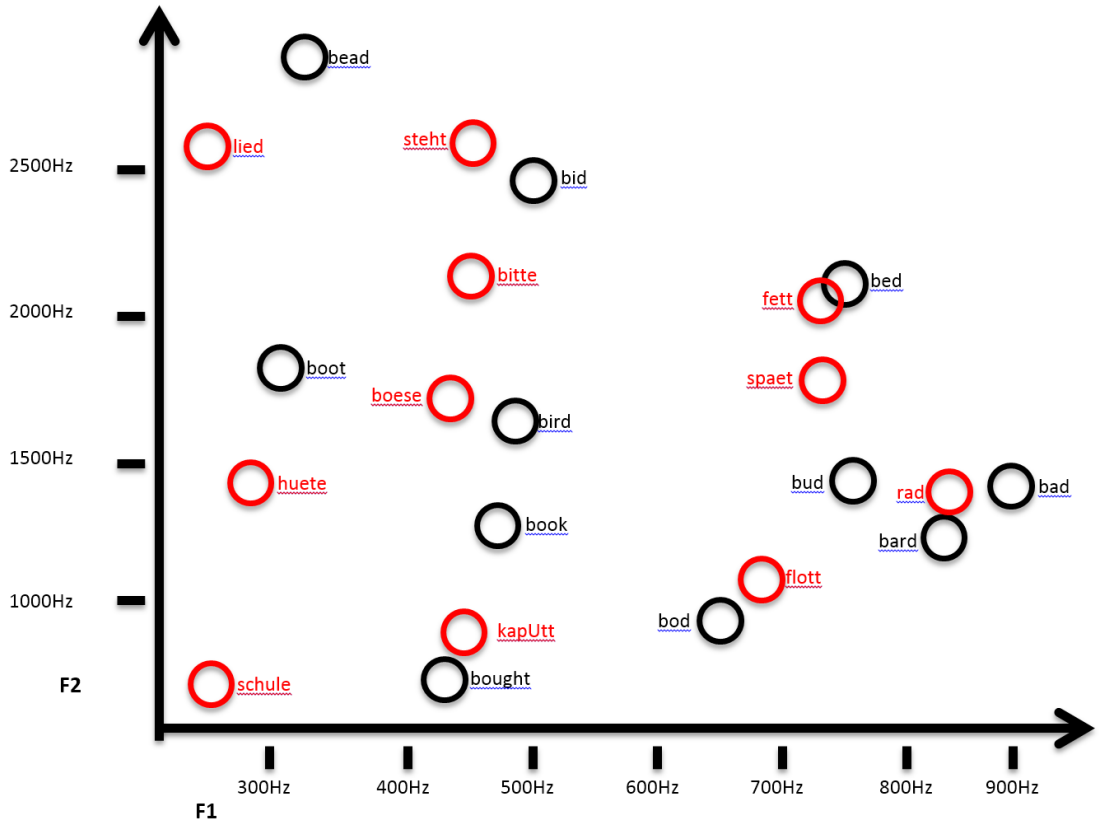

Figure 21: The English and German vowel stimuli (in word context) which were used. The English vowels are depicted in black, the German vowels in red.

Apparatus

The behavioural task was conducted in a sound-proofed room with headphones. A laptop was used to run the experiment and collect responses from the subjects.

Procedure

The English L1 and German L1 subjects performed a combined assimilation and goodness rating task with English vowel categories as targets. Both the L1 and the L2 speakers first heard an English or German vowel and then chose from 11 English target words written on the screen which one contained the vowel they thought

they had heard. The subjects then chose how good a fit the heard vowel was to the chosen target word on a scale of 1 to 5 (5 being a perfect examplar of this vowel). Both assimilation tasks contained English and German vowels as phonemes which needed to be assimilated to either English or German target phonemes. Subjects were not informed that they were hearing English or German vowels. The English L1 speakers performed 110 trials (5 repetitions each of 11 German and 11 English vowels) in 3 blocks, assimilating randomised English and German vowels to English target vowels. The German L1 speakers additionally performed another 110 trials in 3 blocks, assimilating randomised English and German vowels to German target vowels, so the German L1 speakers performed assimilation/goodness tasks for both English and German target vowels.

Results

The subjects were not aware that they were listening to German or English vowels as the vowels were heard on their own, without a word context. Looking at the assimilation data for English L1 speakers (see Table 7), we can see that most English phonemes are categorised correctly by the English L1 speakers. Most of the German vowel sounds are categorised to an English phoneme which is acoustically very close to the German phoneme.

Targets	boot	bead	bid	book	bed	bought	bad	bod	bird	bud	bard
boot	68 (3.51)	0	0	19	1	3	0	1	0	8	0
bead	0	91 (4.09)	7	0	3	0	0	0	0	0	0
bid	0	19	5	0	44 (2.73)	1	17	0	13	0	0
book	11	0	0	28 (3.14)	3	17	0	7	21	13	0
bed	0	3	0	0	68 (3.63)	0	8	0	16	1	4
bought	7	0	0	8	0	61 (3.41)	0	23	0	1	0
bad	0	0	1	0	0	0	63 (3.45)	0	1	1	33 (3.2)
bod	0	1	1	1	0	20	0	41 (3.71)	5	1	28 (3.62)
bird	5	0	0	25 (2.74)	4	5	0	0	48 (3.47)	12	0
bud	0	0	0	0	0	1	19	0	28 (2.86)	7	45 (3.47)
bard	0	0	1	0	0	0	19	0	0	1	79 (3.8)
spaet	0	1	1	0	67 (3.54)	0	5	0	19	0	7
huete	45 (3.06)	7	3	27 (2.85)	1	7	0	0	4	7	0
rad	0	0	1	0	0	0	35 (3.31)	0	3	1	60 (3.47)
schule	72 (4.02)	0	0	15	0	5	0	1	0	7	0
bitte	4	5	8	7	36 (1.93)	3	9	0	23	3	3
lied	0	93 (3.96)	3	0	4	0	0	0	0	0	0
kapUtt	8	0	0	17	0	53 (3.17)	0	21	0	0	0
flott	0	0	0	1	0	19	1	41 (3.58)	5	1	31 (3.74)
boese	20	0	0	16	1	8	1	0	41 (2.87)	12	0
steht	1	21	17	0	37 (2.68)	0	17	0	3	1	1
fett	1	1	1	0	57 (3.23)	0	7	0	25 (3.05)	1	5

Table 7: English L1 speakers. Assimilation of English and German (DE) vowels to English vowel categories. Significant results are highlighted in yellow. Goodness rating in brackets.

Looking at the assimilation data for the L2 speakers with English phonemes as the targets (see Table 8), we can see that assimilating English phonemes to the correct English phoneme target seems to be problematic for the L2 speakers: Only 3 out of 11 assimilations are correct and one is equally divided with another target phoneme.

Targets	boot	bead	bid	book	bed	bought	bad	bod	bird	bud	bard
boot	24	4	8	9	0	9	0	0	35 (1.47)	11	0
bead	0	48 (4.07)	46 (3.03)	0	0	0	0	0	6	0	0
bid	0	11	0	0	72 (3.07)	0	16	0	1	0	0
book	9	1	1	9	1	31 (2.62)	0	12	25	11	0
bed	0	4	0	0	19	0	76 (3.83)	0	0	0	1
bought	5	0	0	7	0	42 (3.47)	0	46 (3.49)	0	0	0
bad	0	0	0	0	0	0	25	0	0	5	71 (3.67)
bod	1	0	0	0	0	64 (3.81)	1	25	0	0	9
bird	6	4	2	6	1	9	0	7	61 (2.75)	4	0
bud	0	5	0	0	0	1	18	0	13	8	55 (3.34)
bard	0	1	0	0	0	0	11	0	0	6	82 (3.91)
spaet	0	8	0	0	22	1	67 (3.39)	0	0	0	1
huete	18	4	2	14	0	13	0	0	39 (1.64)	11	0
rad	0	0	0	0	0	0	9	0	0	5	86 (3.77)
schule	59 (4.08)	0	0	12	0	1	0	0	0	28 (3.29)	0
bitte	4	13	0	2	55 (2.47)	1	4	1	19	1	0
lied	0	45 (4.05)	51 (3.4)	0	0	0	0	0	5	0	0
kapUtt	11	0	0	6	0	31 (3.08)	0	53 (3.33)	0	0	0
flott	4	1	0	1	0	60 (3.88)	2	15	0	0	16
boese	7	5	2	1	1	9	0	1	69 (2.53)	4	0
steht	0	15	0	0	80 (3.06)	0	5	0	0	0	0
fett	0	2	1	0	18	0	73 (3.52)	0	1	1	4

Table 8: L2 speakers. Assimilation of English and German (DE) vowels to English vowel categories. Significant results are highlighted in yellow. Goodness rating in brackets.

When looking at how the L2 speakers assimilated English phonemes and German phonemes to German targets (see Table 9), we can see that the German phonemes are matched up quite well with the German target counterparts.

Targets	spaet	huete	rad	schule	bitte	lied	kapUtt	flott	boese	steht	fett
boot	1	89 (3.89)	0	0	1	0	1	0	2	5	0
bead	0	0	0	1	8	88 (4.13)	2	0	0	0	0
bid	36 (3.1)	0	0	0	0	0	0	0	0	36 (3.81)	27 (2.78)
book	1	1	0	2	5	0	6	34 (1.72)	47 (2.12)	2	1
bed	99 (4.05)	0	0	0	0	0	0	0	0	0	1
bought	0	1	0	0	0	0	0	96 (2.78)	2	0	0
bad	5	0	89 (3.58)	0	0	0	1	0	0	0	5
bod	0	0	22	0	0	0	1	74 (2.59)	2	0	0
bird	0	0	0	0	0	0	0	0	100 (3.53)	0	0
bud	11	0	71 (2.48)	0	0	0	1	1	13	0	4
bard	0	0	99 (3.81)	0	0	0	1	0	0	0	0
spaet	95 (3.67)	0	0	0	0	0	0	0	0	0	5
huete	0	74 (3.54)	0	7	2	2	4	0	1	9	0
rad	0	0	99 (4.13)	0	0	0	1	0	0	0	0
schule	0	4	0	64 (4.11)	0	0	32 (3.56)	0	1	0	0
bitte	15	0	0	0	2	0	0	1	36 (2.61)	34 (3.03)	11
lied	0	0	0	0	6	93 (4.32)	1	0	0	0	0
kapUtt	0	1	0	0	0	0	1	95 (2.67)	2	0	0
flott	1	0	27 (1.39)	1	0	0	0	68 (2.69)	2	0	0
boese	0	0	0	0	2	0	0	1	92 (4.04)	5	0
steht	8	0	0	1	0	0	0	0	1	72 (4.08)	18
fett	98 (3.77)	0	0	0	0	0	0	0	0	0	2

Table 9: L2 speakers. Assimilation of English and German (DE) vowels to German vowel categories. Significant results are highlighted in yellow. The goodness ratings are in brackets.

Discussion

The English L1 speakers seemed to assimilate non-native speech sounds to native vowel categories which were acoustically very close. The German L1 speakers performed at a lower level than the

English speakers when instructed to assimilate English vowels to English vowel categories.

Experiment 2

For the next study, we ran an ACC EEG experiment on all 22 L1 and L2 vowels used in the behavioural tasks in the previous behavioural experiment. The EEG experiment was designed to elicit the ACC response for a change between vowels and to provide data for multidimensional scaling. The aims of this EEG ACC study were: Firstly, to assess the effect of formant change, vowel categories and subject language on the ACC. Secondly, to assess which levels of speech processing are linked to the ACC. Thirdly, to explore what the ACC means in the context of L2 speech sound processing. Fourthly, to attempt multidimensional scaling of the entire vowel space for German and English vowels.

The current study uses the ACC technique in a novel way. Apart from the fact that this study investigates L2 auditory processing with the ACC, which is something only very few studies have done so far (Brint, 2012; Iverson et al., 2013; Martin et al., 2012), there are other aspects which make this study interesting. Previous studies (Brint, 2012; Brown et al., 2008; Friesen & Tremblay, 2006; Kim, Brown, Abbas, Etler, & O'Brien, 2009; Martin & Boothroyd, 2000; Martin et al., 2012; Ostroff et al., 1998) on the ACC have only used 1-2 pairs of sounds. This ACC study uses 155 pairs of German and English vowels. We used this unusually large amount of vowel pairs because we aimed to investigate the whole

vowel space for English and German, not just an extremely limited sub-section, such as previous studies have done. We aimed to do this by using a multidimensional scaling/MDS approach, as recently used successfully on ACC fricative data by Iverson et al. (Iverson et al., 2013). In this study, Iverson et al. (2013) used the magnitude of the ACC response for voiced and voiceless fricative pairs as a similarity measure for MDS. This produced a two-dimensional perceptual space which related to voicing and place of articulation.

Method

Subjects

The subjects for this experiment were English L1 and German L1/English L2 speakers who all participated in the behavioural experiment/Experiment 1. A subset of 19 of these speakers participated in this experiment/Experiment 2. There were nine English Southern British English native speakers (two females and seven males). They were between 23 and 41 years old (mean 30 years old; SD=6.6). The English L2 speakers were ten German native speakers, four females and six males, between 21 and 30 years old (mean 25.8 years old, SD=3, of which seven speakers were from North Germany and three speakers were from South Germany. The German L2 speakers had between 9 and 66 months

of exposure to British English in the UK – mainly living in London or in Southern England (mean 23.6 months residency, SD 21), as well as having learnt English as a foreign language at secondary school.

Stimuli

Stimuli were 11 vowels in English and 11 vowels in German (see Tables 10 and 11), made from British English and Standard German vowels, as used in the preceding behavioural experiment. We created sequences of these vowel stimuli. The duration of each vowel was randomly chosen from 400 to 500ms to jitter durations in order to avoid habituation and therefore flattening of the N1 component (Haenschel, Vernon, Dwivedi, Gruzelier, & Baldeweg, 2005). Stimuli were presented in a random walk of all stimulus pairs and orders. Sequences were presented as long stimulus trains of concatenated and alternating sounds (e.g. "iuiuiuiu…"). The stimulus trains were 350.92s long.

b**oo**t-b**ea**d	b**oo**t-b**i**d	b**oo**t-b**oo**k	b**oo**t-b**e**d	b**oo**t-b**ou**ght	b**oo**t-b**a**d	b**oo**t-b**o**d	b**oo**t-b**i**rd	b**oo**t-b**u**d	b**oo**t-b**a**rd
	b**ea**d-b**i**d	b**ea**d-b**oo**k	b**ea**d-b**e**d	b**ea**d-b**ou**ght	b**ea**d-b**a**d	b**ea**d-b**o**d	b**ea**d-b**i**rd	b**ea**d-b**u**d	b**ea**d-b**a**rd
		b**i**d-b**oo**k	b**i**d-b**e**d	b**i**d-b**ou**ght	b**i**d-b**a**d	b**i**d-b**o**d	b**i**d-b**i**rd	b**i**d-b**u**d	b**i**d-b**a**rd
			b**oo**k-b**e**d	b**oo**k-b**ou**ght	b**oo**k-b**a**d	b**oo**k-b**o**d	b**oo**k-b**i**rd	b**oo**k-b**u**d	b**oo**k-b**a**rd
				b**e**d-b**ou**ght	b**e**d-b**a**d	b**e**d-b**o**d	b**e**d-b**i**rd	b**e**d-b**u**d	b**e**d-b**a**rd
					b**ou**ght-b**a**d	b**ou**ght-b**o**d	b**ou**ght-b**i**rd	b**ou**ght-b**u**d	b**ou**ght-b**a**rd
						b**a**d-b**o**d	b**a**d-b**i**rd	b**a**d-b**u**d	b**a**d-b**a**rd
							b**o**d-b**i**rd	b**o**d-b**u**d	b**o**d-b**a**rd
								b**i**rd-b**u**d	b**i**rd-b**a**rd
									b**u**d-b**a**rd

Table 10: English word pairs (bold=target vowels)

sp**ae**t-h**ue**te	sp**ae**t-r**a**d	sp**ae**t-sch**u**le	sp**ae**t-b**i**tte	sp**ae**t-l**ie**d	sp**ae**t-kap**u**tt	sp**ae**t-fl**o**tt	sp**ae**t-b**oe**se	sp**ae**t-st**e**ht	sp**ae**t-f**e**tt
	h**ue**te-r**a**d	h**ue**te-sch**u**le	h**ue**te-b**i**tte	h**ue**te-l**ie**d	h**ue**te-kap**u**tt	h**ue**te-fl**o**tt	h**ue**te-b**oe**se	h**ue**te-st**e**ht	h**ue**te-f**e**tt
		r**a**d-sch**u**le	r**a**d-b**i**tte	r**a**d-l**ie**d	r**a**d-kap**u**tt	r**a**d-fl**o**tt	r**a**d-b**oe**se	r**a**d-st**e**ht	r**a**d-f**e**tt
			sch**u**le-b**i**tte	sch**u**le-l**ie**d	sch**u**le-kap**u**tt	sch**u**le-fl**o**tt	sch**u**le-b**oe**se	sch**u**le-st**e**ht	sch**u**le-f**e**tt
				b**i**tte-l**ie**d	b**i**tte-kap**u**tt	b**i**tte-fl**o**tt	b**i**tte-b**oe**se	b**i**tte-st**e**ht	b**i**tte-f**e**tt
					l**ie**d-kap**u**tt	l**ie**d-fl**o**tt	l**ie**d-b**oe**se	l**ie**d-st**e**ht	l**ie**d-f**e**tt
						kap**u**tt-fl**o**tt	kap**u**tt-b**oe**se	kap**u**tt-st**e**ht	kap**u**tt-f**e**tt
							fl**o**tt-b**oe**se	fl**o**tt-st**e**ht	fl**o**tt-f**e**tt
								b**oe**se-st**e**ht	b**oe**se-f**e**tt
									st**e**ht-f**e**tt

Table 11: German word pairs (bold=target vowels)

Apparatus

A Biosemi EEG system with 64 electrodes (secured using a cap and electrolyte gel) was used for both EEG experiments. Seven additional electrodes were used: The left mastoid, the right mastoid, the left canthus, above the left eye, the cheek bone (below the left eye), the right canthus and the bridge of the nose. The nasal bridge electrode was used as earth. The cap was positioned on the head ensuring that Cz was at the vertex of the head. This was done by placing it half way between the subject's nasion and inion and half way between the bilateral pre-auricular points. See Figure 23 below for the placement of the electrodes.

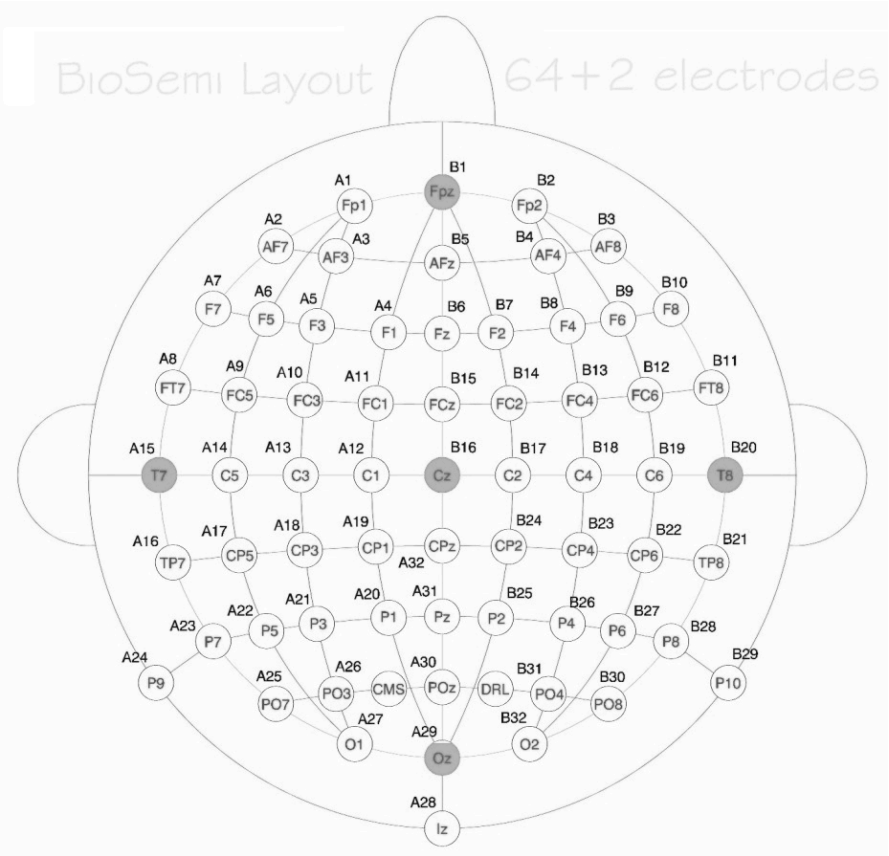

Figure 22: Biosemi EEG system electrode layout

The sound was presented through an external sound-card via insert headphones (Etymotics ER3) and subjects were tested in a magnetically shielded sound proof booth which was kept at 19 degrees celsius in order to avoid sweat artefacts. The visual task was presented on a laptop (run on batteries) placed directly in front of the subject. Mains electricity in the booth was switched off in order to avoid artefacts.

Procedure

The total duration of the ACC EEG experiment was 90 minutes (plus set-up time for attaching electrodes). During playback of the stimuli over insert headphones the subjects watched nonstop Tom & Jerry cartoon episodes (without sound). Subjects listened to 12 blocks of stimuli and stimuli were blocked by language (English or German), resulting in 84 epochs/vowel pair. Each block took 5.7 minutes and subjects were offered a break after every block in order to prevent fatigue. The stimuli were presented at 60 dB/sound pressure level. Subjects were asked to remain as calm and relaxed as possible while watching the film, to try and ignore the sounds played over the headphones and to try not to blink too often.

Pre-processing

Using SPM in Matlab, all the EEG data went through the following pre-processing steps: The data was montaged to reference the

electrode channels to the nose channel. Then the CEOG and VEOG channels were calculated. The EEG data was filtered twice, using a high-pass Butterworth filter (1.15 Hz) and a low-pass Butterworth filter (30 Hz) and downsampled to 128 Hz. Eyeblink and muscle artefacts were rejected using Independent Component Analysis (ICA) with EEGlab. The data was then corrected for channels that seemed related to eye artefacts. After this, the data was epoched, the epochs being from 100ms before to 350ms after stimulus change, and baseline-corrected (using the 100ms before stimulus change as a baseline). Then, artefact rejection with a threshold of 100 microvolts was applied. Data were then averaged across all epochs for each subject. One subject was excluded from further data analysis due to a very high amount of eye-blinks. The total amount of epochs per stimulus language was 9240 (220 vowel pairs x 42). More than 90% of the trials were usable.

Analysis

1) Methods for measuring the ACC

The average response (across all subjects and stimulus changes) is plotted below (see Figure 24). The largest amplitude ACC response occurs at the central electrode locations, with the response changing in morphology toward the back and sides. Conventionally, one would measure the ACC as the magnitude of the response at

one of the central electrodes (e.g., FCz or Cz). However, according to Näätänen & Picton (Näätänen & Picton, 1987), the temporal electrodes (T7 and T8) may show a different morphology to the rest of the scalp. This component of the ACC is called the T-complex and consists of the components P1 and N150 at electrode locations T7 and T8. We were interested in seeing if we could reproduce the T-complex with our data, but decided to attempt this in a more data driven way, rather than by looking at the data visually. Figure 25 shows the average ACC response over all subjects and stimuli.

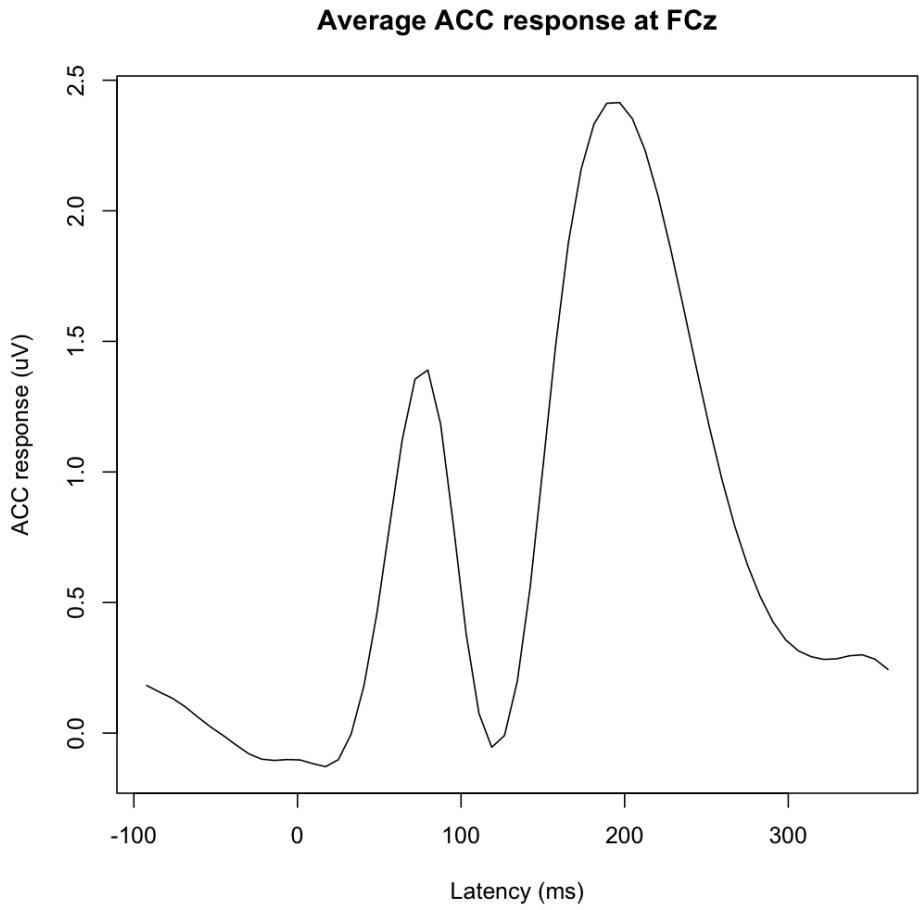

Figure 23: ACC average response at FCz (across language groups and all stimuli)

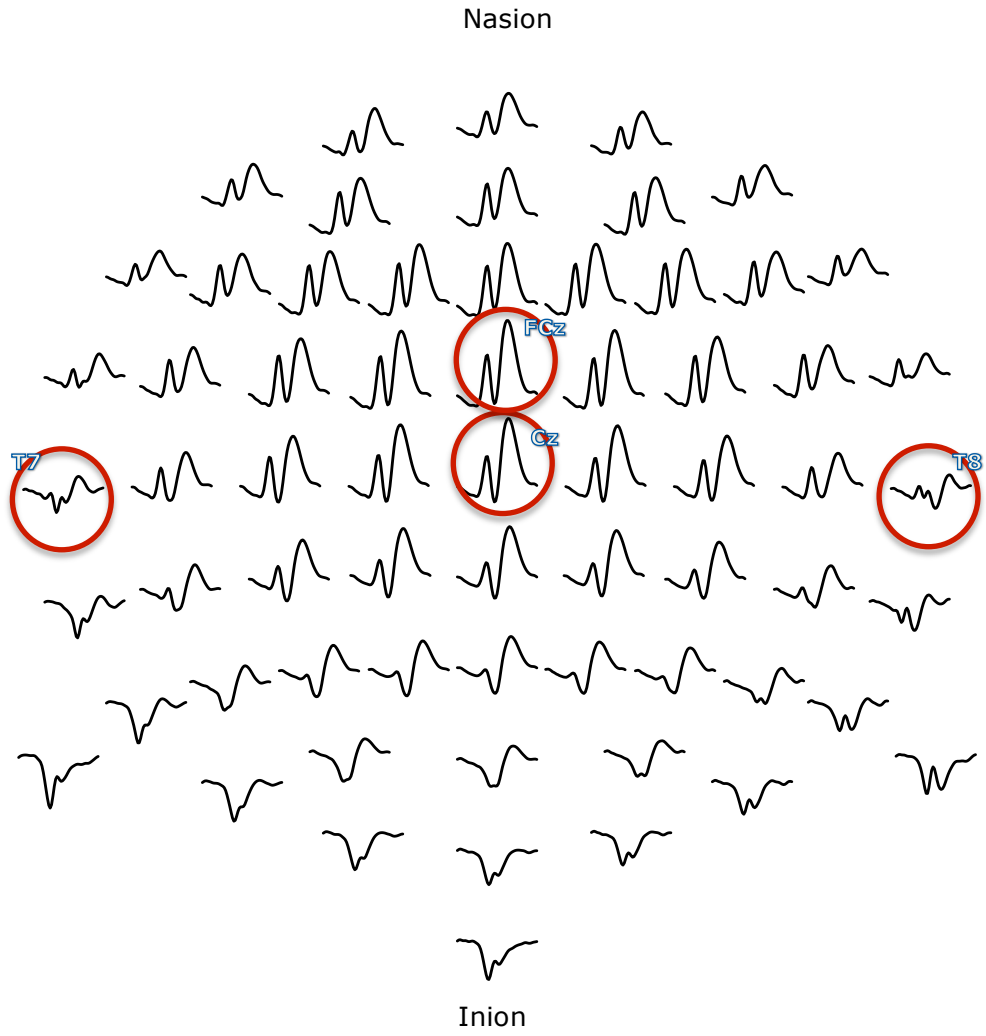

Figure 24: Average ACC across all subjects and stimulus changes. FCz and Cz are conventionally used to plot ACC responses. However, the temporal electrodes T7 and T8 can show a different morphology to the midpoint electrodes. Positive values are plotted upwards

We used cluster analysis in an attempt to make this choice more data driven. We averaged data across all subjects, and used the variability across stimulus pairs to group the electrodes. That is, two electrode sites would tend to group together if the response became greater or lesser for the same stimulus pairs. Below (see

Figure 25) is the cluster analysis based on response magnitude (root mean square/RMS across the ACC response).

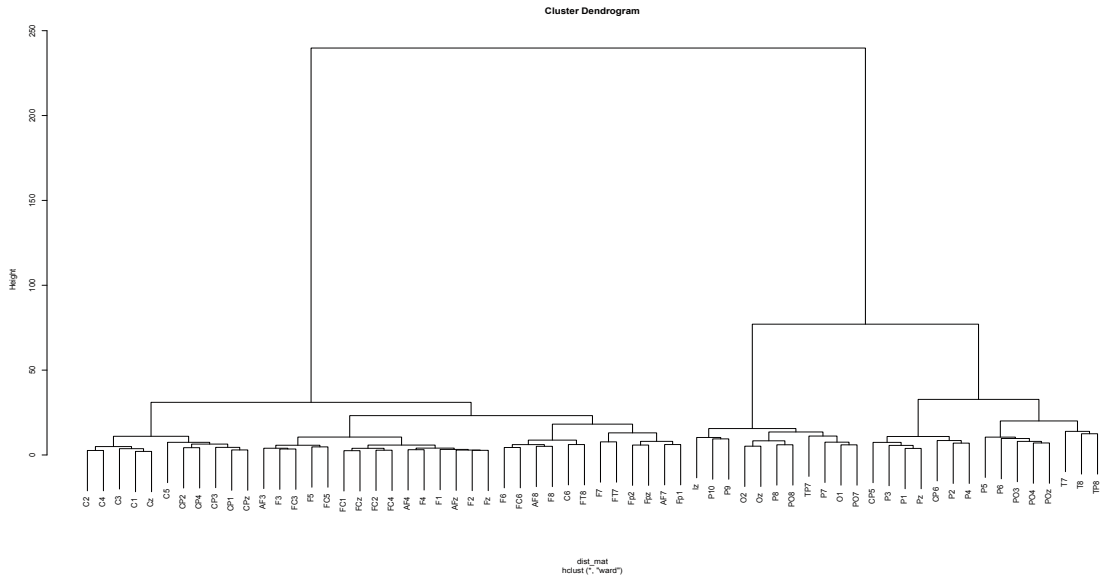

Figure 25: Cluster analysis of the ACC data for all subjects. The clusters are based on RMS.

The results demonstrated that there are two very distinct sets of electrodes, with the left-most group possibly divisible into two. To display these results, the average ACC responses with the two clusters (the frontocentral and temporoparietal regions) are plotted (see Figure 26).

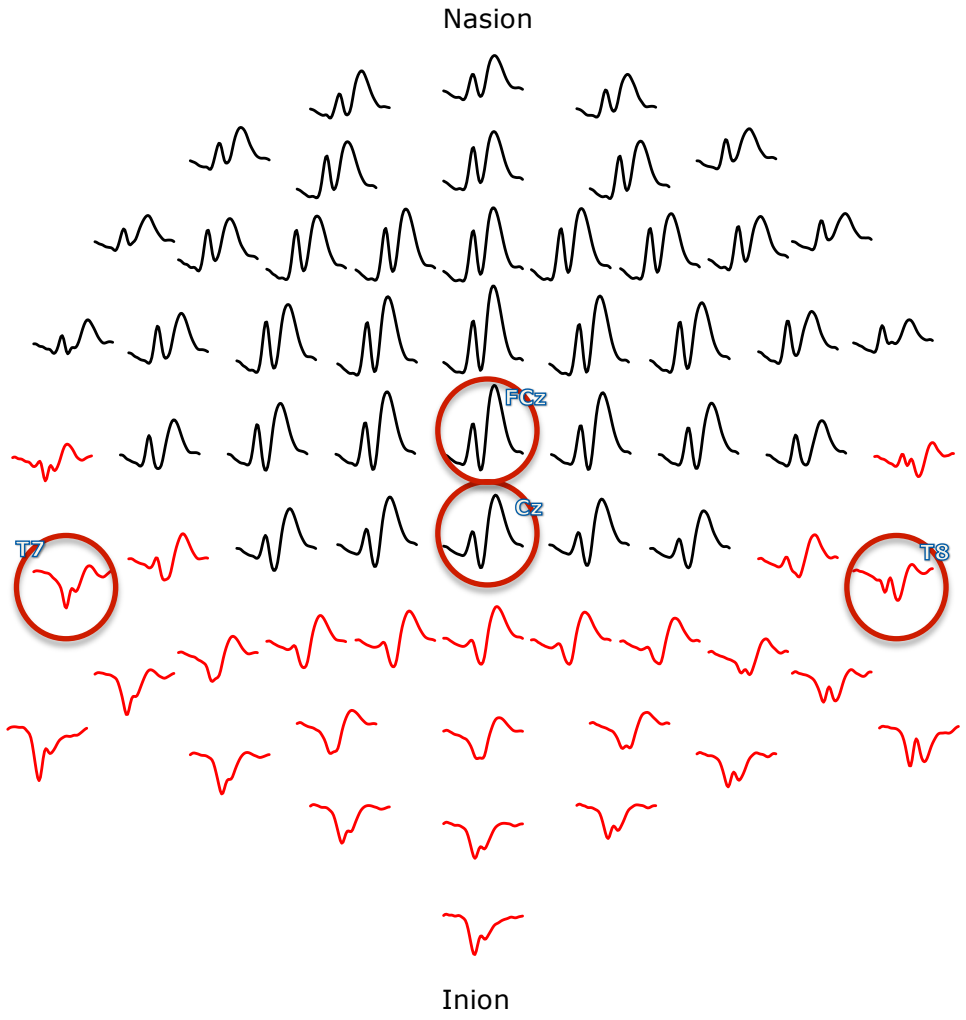

Figure 26: Average responses for all subjects based on RMS. The frontocentral cluster is displayed in black and the temporoparietal cluster in red.

The temporoparietal sites give an ACC response that is very distinct from the main frontocentral response. This is equivalent to the T-complex with the onset P1-N1-P2 complex. Below (see Figure 27) is a similar cluster analysis, except this time it is based on N1 latency. To display these results, the average ACC responses based on N1 latency are plotted below with the two clusters again plotted in red and black (see Figure 28).

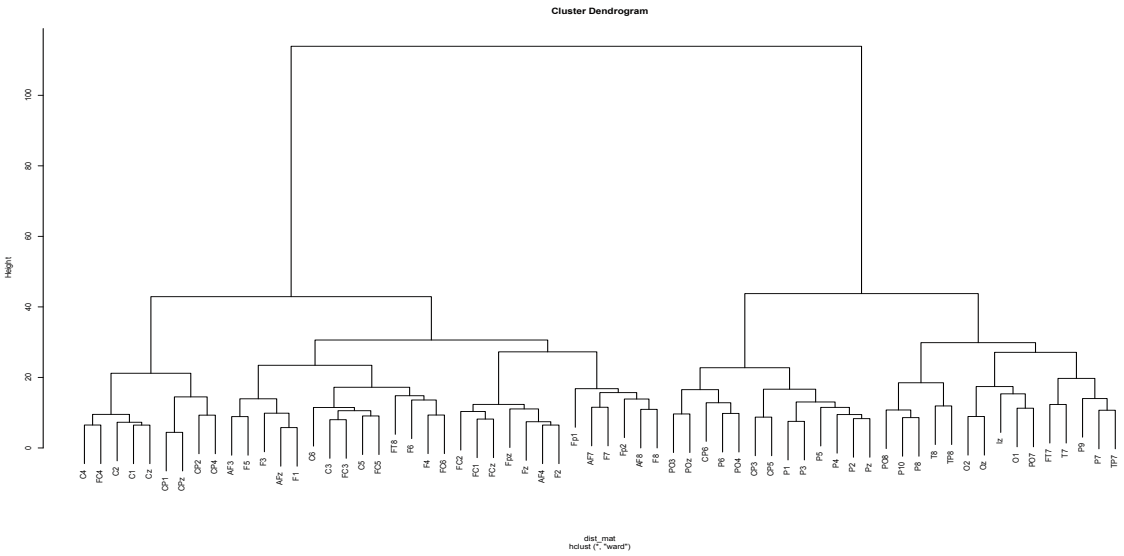

Figure 27: Cluster analysis of the ACC data for all subjects. The clusters are based on N1 latency.

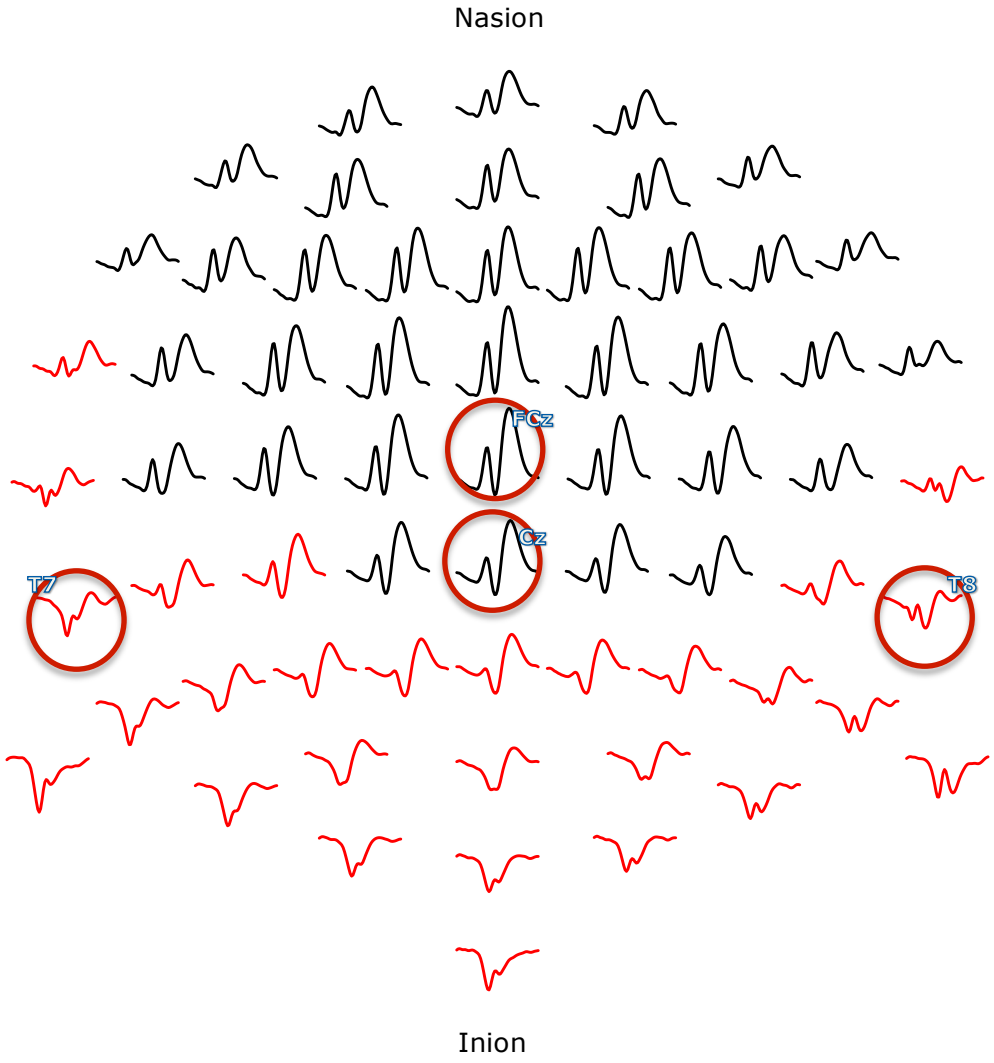

Figure 28: Average ACC responses for all subjects, with frontocentral and temporoparietal clusters in black and red respectively. The clusters are based on N1 latency.

The cluster analysis based on N1 latency is very similar to that based on magnitude (root mean square/RMS), with the clusters being not quite as symmetric with regard to laterality (i.e., there are two more electrodes in the left hemisphere joining the temporal/parietal group).

Based on these analyses, we decided to measure the ACC in

four ways, based on localisation (ie. the two different clusters of responses/electrode sites; frontocentral and temporoparietal), and in terms of overall magnitude (RMS) and N1 latency.

2) Plots of responses for matrices

The FCz and P9 electrodes were identified as the largest responses in their respective cluster and were thus chosen for analysis (see Figure 30).

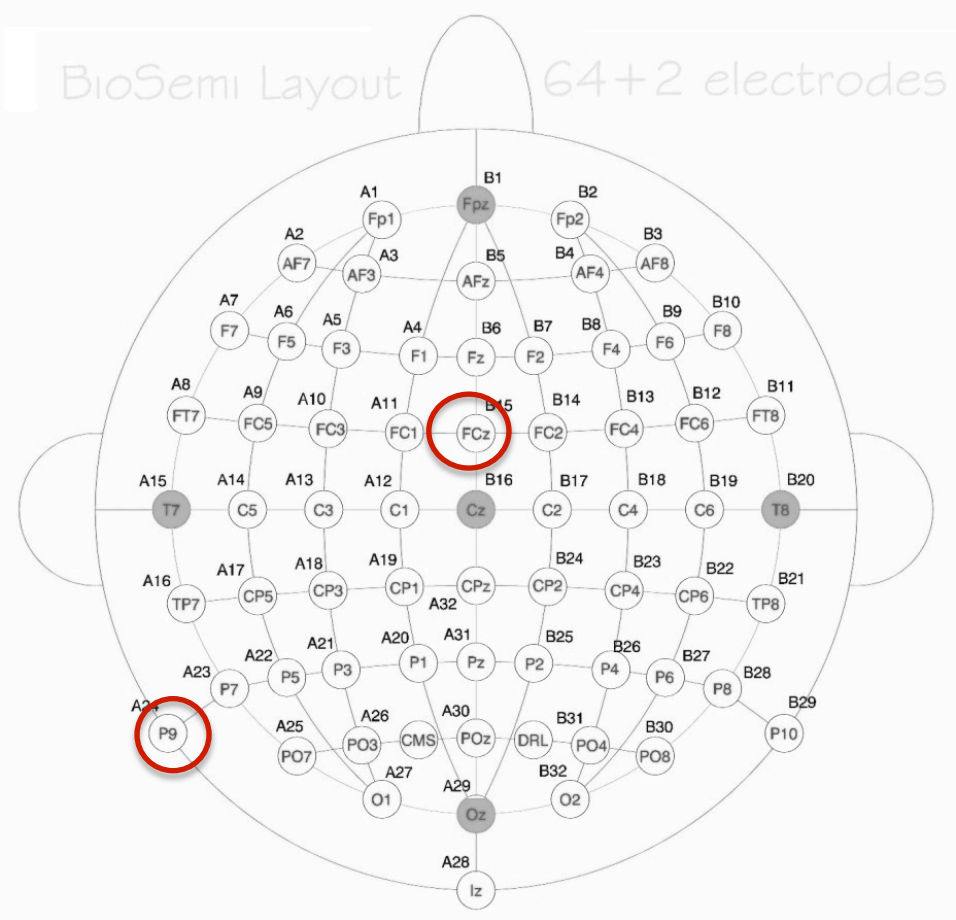

Figure 29: Biosemi diagram of scalp electrodes, FCz and P9

Below is the response matrix at FCz (which was the largest response of the frontocentral cluster) for the English vowel pairs (see Figure 31) and the German vowel pairs (see Figure 32), with black for English listeners and red for German listeners, positive-going waves plotted upwards.

When looking at the English vowel pair matrix, it is immediately visible that certain vowel pairs show more of a cross-language difference between the two groups of speakers than others, specifically “bid-bud”, “boot-bud”, “bid-bird”, “boot-bird”, “bid-bod”, “book-bought”, “book-bird”, “bed-bought”, “bed-bard”, “bed-bad”, “bought-bird”, “bod-bird”, “bod-bud” and “bird-bard”.

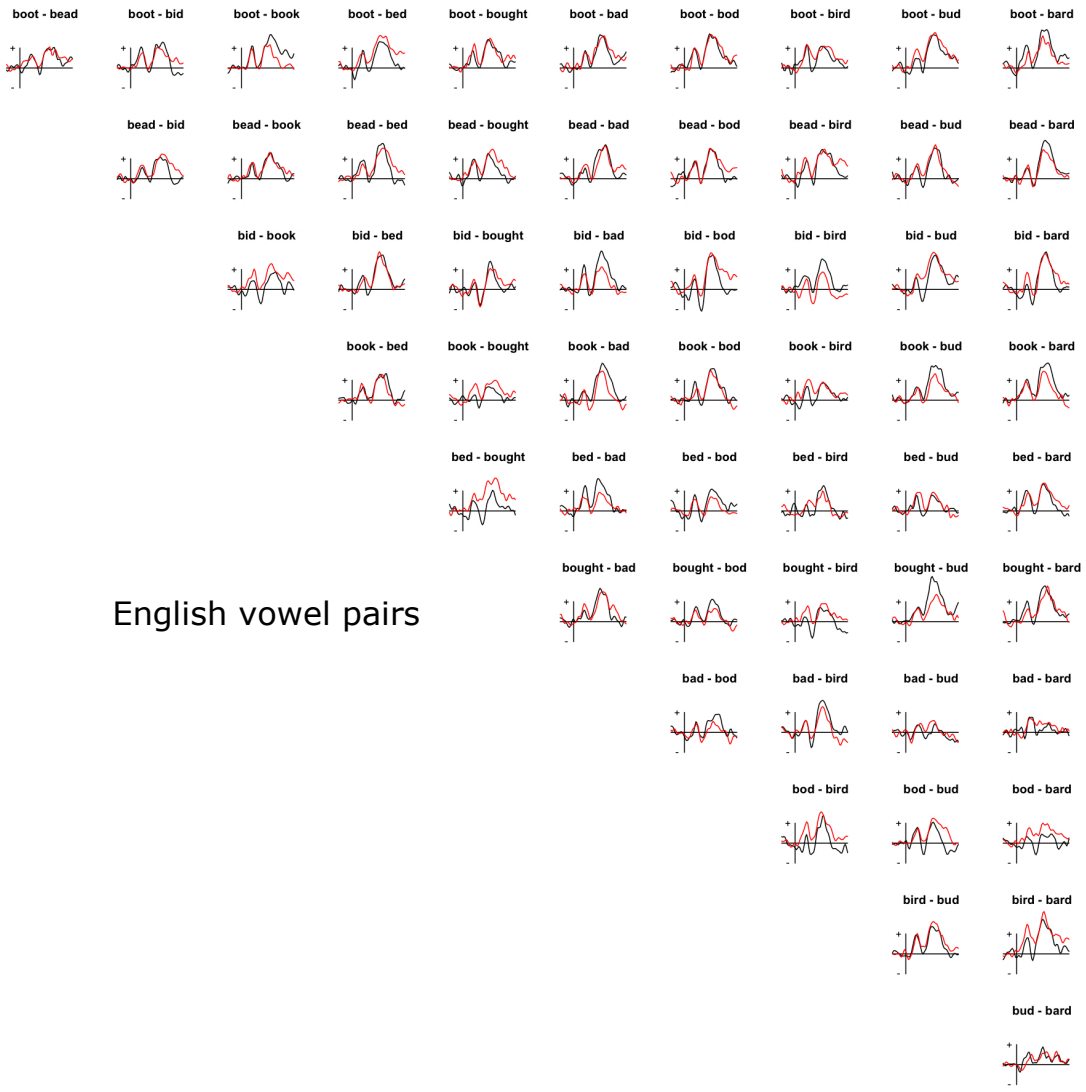

Figure 30: English vowel pairs ACC at FCz (black = English L1 speakers and red = German L1 speakers), positive plotted upwards. Range of +/- 2 microvolts. The y-axis crosses the x-axis at the stimulus change (100 ms pre-stimulus interval, followed by 350 ms post stimulus).

The German vowel pair matrix (see Figure 32) shows visible cross-language differences between the two groups of speakers as well, specifically in the "spaet-bitte", "spaet-rad", "huete-steht", "rad-steht", "schule-kapUtt", "bitte-flott", "bitte-boese", "lied-steht",

"kapUtt-boese", "kapUtt-fett" and "flott-fett" German vowel pairs.

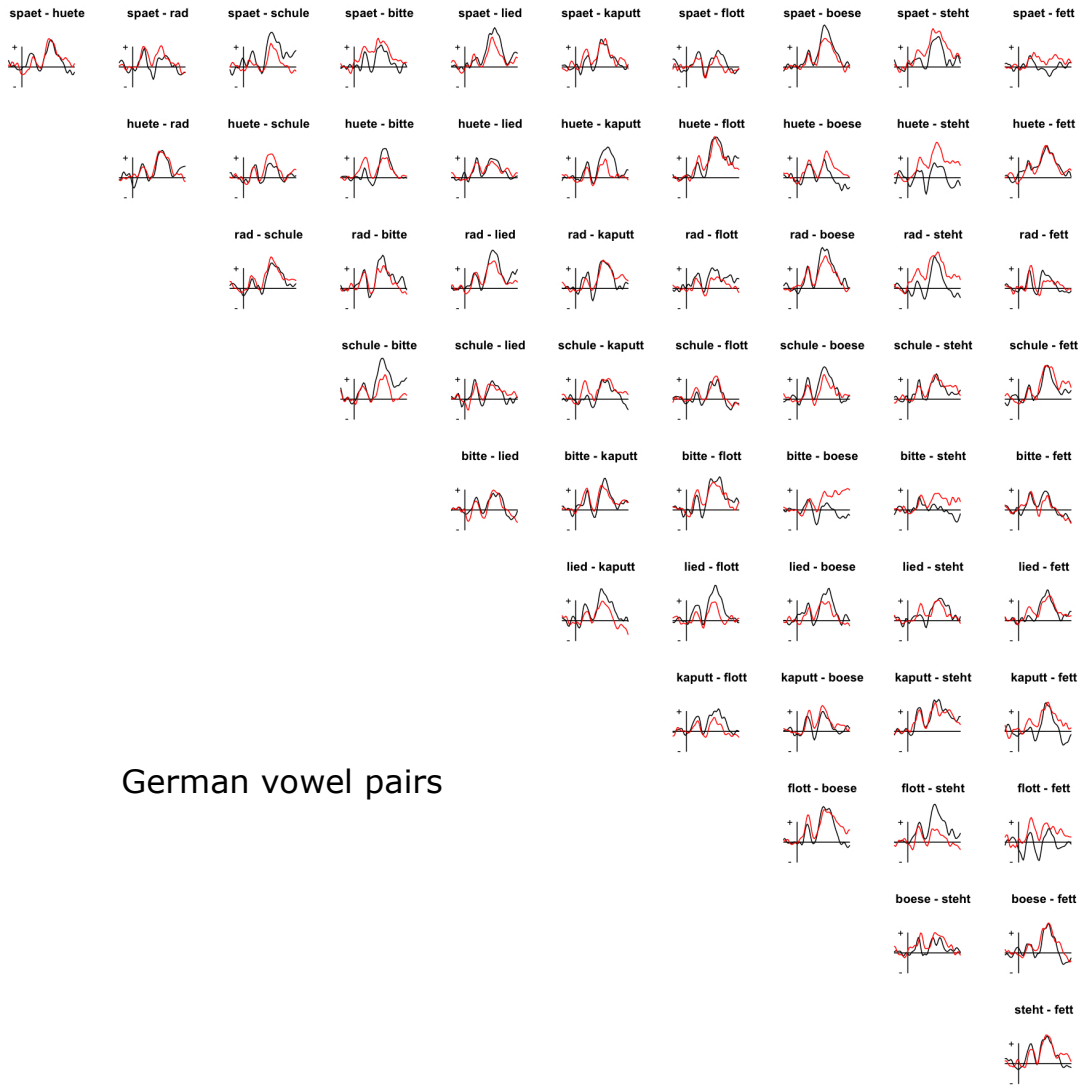

Figure 31: German vowel pairs ACC at FCz (black = English L1 speakers and red = German L1 speakers), positive plotted upwards. Range of +/- 2 microvolts. The y-axis crosses the x-axis at the stimulus change (100 ms pre-stimulus interval, followed by 350 ms post stimulus).

Below is a similar display at the electrode location P9, which was the electrode with the largest average response in the

temporoparietal cluster (see Figure 33 for English vowel pairs and Figure 34 for German vowel pairs). What is immediately obvious is that the ACC responses of the temporoparietal region look different to the ACC responses obtained from the frontocentral region. The P1 and P2 components are not as pronounced in this region. When looking at the only reliably present component N1, the English vowel pairs "boot-bid", "boot-bed", "boot-bad", "boot-bird", "boot-bard", "beat-bard", "bead-bod", "bid-bed", "bid-bard", "book-bed", "book-bought", "book-bird", "bed-bod", "bought-bird", "bod-bud", "bod-bard", "bird-bard" and "bud-bard" show a cross-language difference between the two groups of speakers. A visible cross-language difference with the English vowel pairs "boot-bird", "bought-bird", "bod-bud" and "bird-bard" is common to both frontocentral and temporoparietal regions.

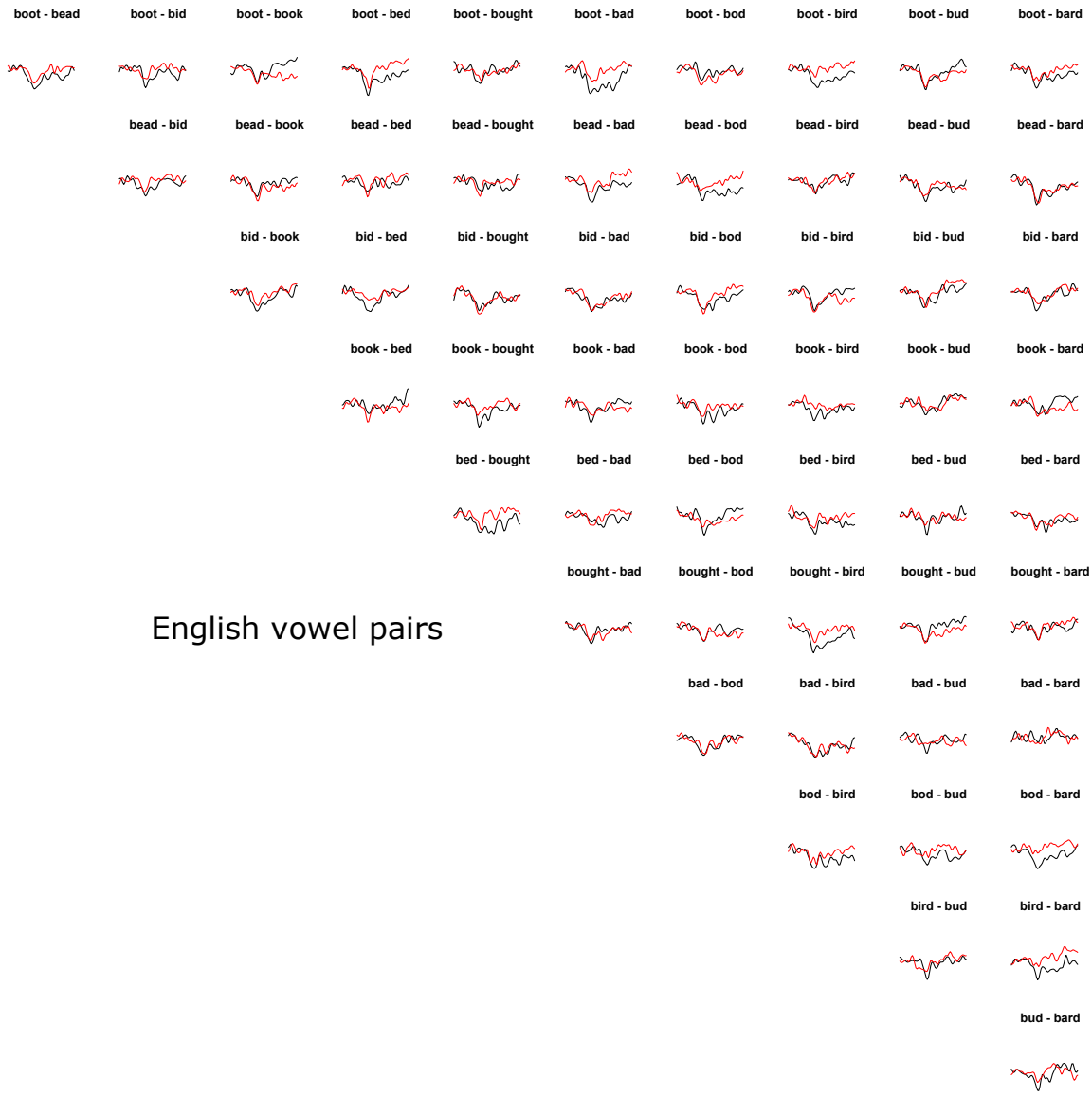

Figure 32: English vowel pairs ACC at P9 (black = English L1 speakers and red = German L1 speakers), positive plotted upwards.

Looking at component N1, the German vowel pairs "spaet-schule", "spaet-lied", "spaet-kapUtt", "spaet-flott", "spaet-steht", "huete-bitte", "huete-lied", "huete-kapUtt", "huete-flott", "huete-boese", "huete-steht", "schule-lied", "schule-steht", "bitte-boese", "bitte-steht", "kapUtt-boese", "flott-boese", "flott-fett" and "steht-fett"

show a visible cross-language difference between the two groups of speakers. What stands out in this response data is that the vowel pairs which involve a German umlaut (specifically "boese" and "huete") seem to elicit larger cross-language differences between the two groups of speakers than other vowels.

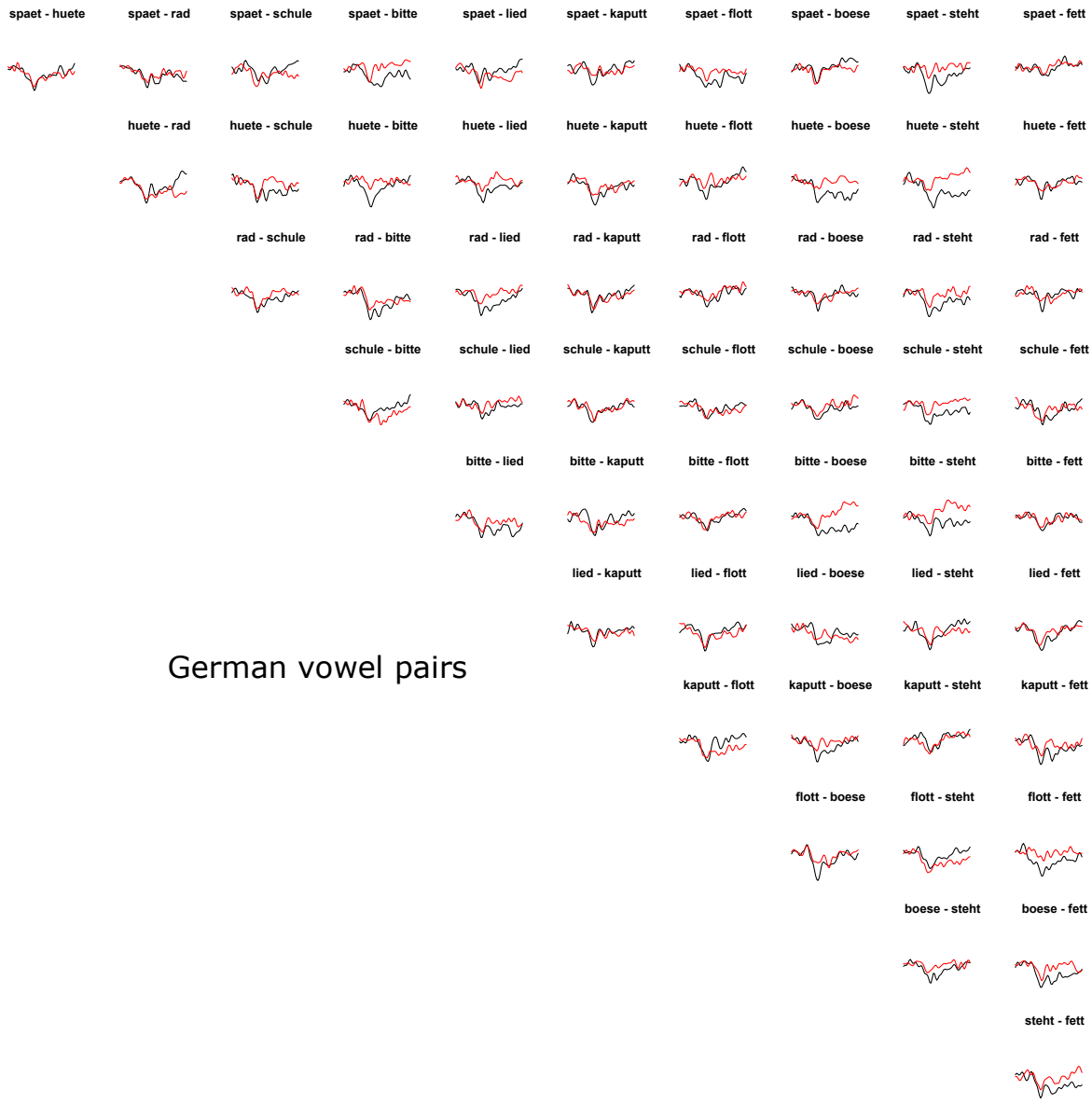

Figure 33: German vowel pairs ACC at P9 (black = English L1 speakers and red = German L1 speakers), positive plotted upwards.

3) Mixed-model analyses

We decided to analyse the relationships between the ACC response data and the behavioural data using mixed-model analysis due to the fact that we had subject groups of different sizes. This is not a problem for mixed-model analysis.

The English and German vowel pair matrices were analysed with four separate mixed-model analyses (RMS magnitude or N1 latency, frontocentral cluster or temporoparietal cluster). For each analysis, the scores were entered for individual subjects and all of the pairs, using the highest-amplitude electrode site within the cluster (FCz for the frontocentral cluster and P9 for the temporoparietal cluster).

Frontocentral RMS magnitude

We started by investigating the response magnitude of FCz in the frontocentral cluster. To start with, we ran a very basic mixed model analysis with magnitude fit with the fixed factors of stimulus pair, subject language (English or German) and vowel language (English or German), and subject as a random factor. There was a significant difference between stimulus pairs only ($F(1,366)=5.2$, $p<0.05$).

However, we wanted to analyse differences between vowel pairs, so we next moved vowel pair to being a random effect and put formant difference as a predictor. The AIC criteria of the second

model went up (indicating not as good a fit), but BIC went down to the fewer degrees of freedom. It is clear, though, that formant difference is a strong predictor ($F(1,3437)=227.7$, $p<0.05$), and we almost get a significant difference with subject language, $F(1,17)=3.7$, $p=0.054$.

However, there is a sense in which the categorical difference between the languages of the stimuli and listeners is too crude, when in reality the stimuli varied in terms of the degree in which they sounded more like German or English to the listeners (i.e., listeners were not told explicitly which they were hearing). Stimuli also varied in terms of how closely they assimilated into a single L1 category. So we replaced subject and vowel language with a measure of L1 category goodness (i.e., the average goodness rating for a pair of vowels in terms of the listener's native language) and L1 category assimilation, which is a normalized chi-square measure of whether the two vowels received the same or different categorization judgements; see (Iverson, Bernstein, & Auer, 1998). The goodness and assimilation measures were those acquired in Experiment 1. L1 category assimilation has a significant effect ($F(1,3432)=24.65$, $p<0.001$) in this model, as does formant frequency ($F(1,3432)=230.37$, $p<0.001$). When we did this the model improved (the AIC criteria went down, indicating a better fit), strongly indicating that there is a role of categorization.

The next step is to see whether this individual effect of categorization is due to a real cross-language difference rather than something more idiosyncratic. We swapped the individual assimilation and goodness scores for group L1 scores (i.e., the scores for English subjects calculated across the ratings for all of the subjects). L1 assimilation and L1 goodness scores were significant ($F(1, 3432)=53.46$, $p<0.001$) and $F(1,3432)=41.46$, $p<0.001$), respectively). This improved the fit for the model further (AIC criteria decreasing further, indicating a better fit), indicating that the group measures were a little less variable.

However, with this model we do not really know whether the specific L1 of the listener matters, so we tried entering both the English and German assimilation ($F(1,3408)=33.11$, $p<0.001$ and $F(1,3408)=36.77$, $p<0.001$, respectively) and English and German goodness ($F(1,3408)=41.66$, $p<0.001$ and $F(1,3408)=40.72$, $p<0.001$, respectively) into the model, regardless of whether the listener was English or German. It significantly improved the model fit (the AIC criteria decreasing further), which undermines the hypothesis that cross-language differences matter.

After this, though, we tried to enter subject language back in, given that this factor is no longer in any of the predictor variables. The model improves (i.e. further reduction in AIC criteria), indicating that we do indeed have significant cross-language

differences (i.e. an interaction of formant frequency and subject language ($F(1,3377)=4.08$, $p<0.05$). Unfortunately, they get buried within a complex statistical model. Finally, we tried to enter individual differences in identification and discrimination into the model, but this step did not improve the results.

Frontocentral N1 latency

The next set of analyses investigated the N1 response latency at the frontocentral cluster. The latency analysis was very similar to the previous analysis. The same factors were significant, but, additionally, individual differences in English ID abilities also contributed to the model.

Temporoparietal RMS magnitude

The next set of analyses focused on the response magnitude in the temporoparietal cluster. We found the best fitting model, following the same strategy as above. What is most notable about this model is that there is no effect of formant difference; the magnitude of the acoustic change does not appear to change the response at this site. There are effects of categorization, but only individual ones ($F(1,3436)=3.88$, $p<0.05$); there is no evidence of categorization effects related to whether the individual is English or German.

Temporoparietal N1 latency

The next set of analyses investigated N1 response latency in the temporoparietal cluster. The latency analysis were similar in that

formant frequency did not play a role, but the effects of goodness, identification, and subject language were similar to that for the frontocentral sensors.

To summarise, for the frontocentral magnitude measured at FCz, there were effects of categorisation, goodness, assimilation and subject language. There was also an effect of acoustic difference between stimuli. There was, however, no effect of ID abilities. For the frontocentral N1 latency measured at FCz, there were effects of categorisation, goodness, assimilation, acoustic difference, subject language and an effect of individual differences in English ID abilities. There was however, no effect of general ID abilities.

For the temporoparietal magnitude measured at P9, there were individual effects of categorisation, but there was no effect of acoustic difference. For the temporoparietal N1 latency measured at P9, there were effects of goodness, assimilation and subject language. As with the magnitude, latency also showed no effect of acoustic difference.

4) Multidimensional scaling

Due to our efforts to cover the whole vowel space for German and English and to not fatigue and demotivate our participants too much we did not have the necessary amount of trials/condition to run an

MDS analysis on the data. We had originally planned to run more blocks with the same stimuli in order to acquire enough trials for MDS analysis, but this proved impossible as it made the experiment too long for the participants and they could not concentrate any longer. As MDS was not a main focus of this experiment we decided to keep the experimental design as is and forego MDS as the paradigm and stimuli worked for ACC.

Experiment 3

We decided to shorten the time spent recording EEG to avoid subjects becoming too fatigued. This meant that we had to either concentrate on fewer vowel stimuli or run many vowel stimuli with fewer trials/vowel.

The previous study showed that 80 trials/vowel change was not a sufficient number of trials to obtain a clean ACC response or to perform MDS on our data. The previous study also showed that while the effects of categorisation were significant they ended up being embedded in complicated statistical analyses. The cross-linguistic effects were there, but they were spread too thin, due to our design having too many vowel pairs and not enough repetitions. Iverson et al. (Iverson et al., 2013) were successful with MDS based on a smaller number of fricatives. So we thought we could use MDS too on a broader vowel space based on fewer repetitions and with more of a structure. All of these considerations, based on the results from the previous experiment, led to us designing a new ACC experiment (the current study) which combines a high number of trials/vowel, while at the same time keeping total EEG recording time lower than previously. A cleaner ACC response would mean we could analyse more aspects of the EEG response.

Our results from the previous experiment showed evidence of cross-language differences, but the data was not good enough to

perform MDS. We decided to concentrate on only a few maximally diverging vowel stimuli for each language and to run them with many more repetitions in order to get a cleaner ACC response. In order to choose our vowels, we looked at the results from Experiment 2 and chose vowel candidates where the ACC response diverged visually the most between L1 and L2 speakers. We chose our vowel candidates for this experiment by looking at the ACC data from the previous experiment (Experiment 2, Figures 32 and 33) and choosing pairs which visually showed maximum divergence between L1 and L2 speakers for the ACC.

Our new experimental design is not suitable for the MDS approach as we have too few vowel pairs. However, in order to get a clearer view of what is happening with respect to the cross-linguistic differences, we thought it would be safer to focus on a particular area of the vowel space/language which seemed to be causing problems for each language group and to give the subjects many more trials. This approach was expected to give us cleaner ACC data for analysis.

The aims of this second ACC EEG study were firstly, to assess the dependence of the ACC on changes in stimulus formants and subject language. Secondly, to obtain more stable ACC data through a larger number of trials/condition in order to be able to analyse all three components of the ACC separately. Thirdly, to

focus on a subset of all vowels which seemed to be problematic for the English or German speakers following visual inspection of data from the first EEG experiment.

Method

Subjects

All subjects from the first EEG experiment took part in the second EEG experiment, the current study. Additional subjects were recruited for the second experiment. The subjects from the behavioural tasks (Experiment 1) and the current, second, EEG experiment were identical: Fifteen Southern British English L1 subjects and sixteen German L1 subjects. One L2 speaker was excluded from EEG data analysis due to excessive blinking artefacts.

Stimuli

Stimuli were a subset of four English and four German vowels from the previous ACC EEG experiment, six pairs per language (see Table 12). They were identical to the stimuli used in the previous EEG experiment. We concentrated on a subset of English and German vowels which had showed visible cross-language differences in the previous ACC experiment.

b**u**d-b**e**d	b**u**d-b**a**d	b**u**d-b**o**d
	b**e**d-b**a**d	b**e**d-b**o**d
		b**a**d-b**o**d

b**oe**se-h**ue**te	b**oe**se-st**e**ht	b**oe**se-sch**u**le
	h**ue**te-st**e**ht	h**ue**te-sch**u**le
		st**e**ht-sch**u**le

Table 12: English and German vowel pairs (bold=target vowels). Stimuli for this second ACC experiment were a subset of the vowel stimuli for the first ACC experiment. Stimuli were chosen which had shown cross-language differences in the first ACC experiment.

Apparatus

The EEG experimental setup was identical to the equipment used in the previous experiment.

Procedure

The procedure was mostly the same as for Experiment 2. The Tom & Jerry cartoons shown were different to the ones from Experiment 2. EEG recording took 45 minutes and the subjects listened to 8 blocks of stimuli. Stimuli were not blocked by language for this experiment. Each block took 5.5 minutes. This resulted in 448 epochs/vowel pair in total per subject.

Pre-processing

The data from EEG experient 3 was pre-processed identically to the data from EEG experient 2.

Analysis

1) Matrices at FCz

Below is the response matrix at FCz (which was the largest response of the frontocentral cluster, as established by Experiment 2) for the English vowel pairs (see Figure 35) and the German vowel pairs (see Figure 36), with black for English listeners and red for German listeners, positive-going waves plotted upwards. When looking at the English vowel pair matrix, it is immediately visible that all vowel pairs show a clear cross-language difference between the two groups of speakers for N1. The German vowel pair matrix shows a less pronounced, but present, cross-language difference for N1, however, the P2 cross-language difference seems more pronounced, especially for the vowel pair “boese-steht”.

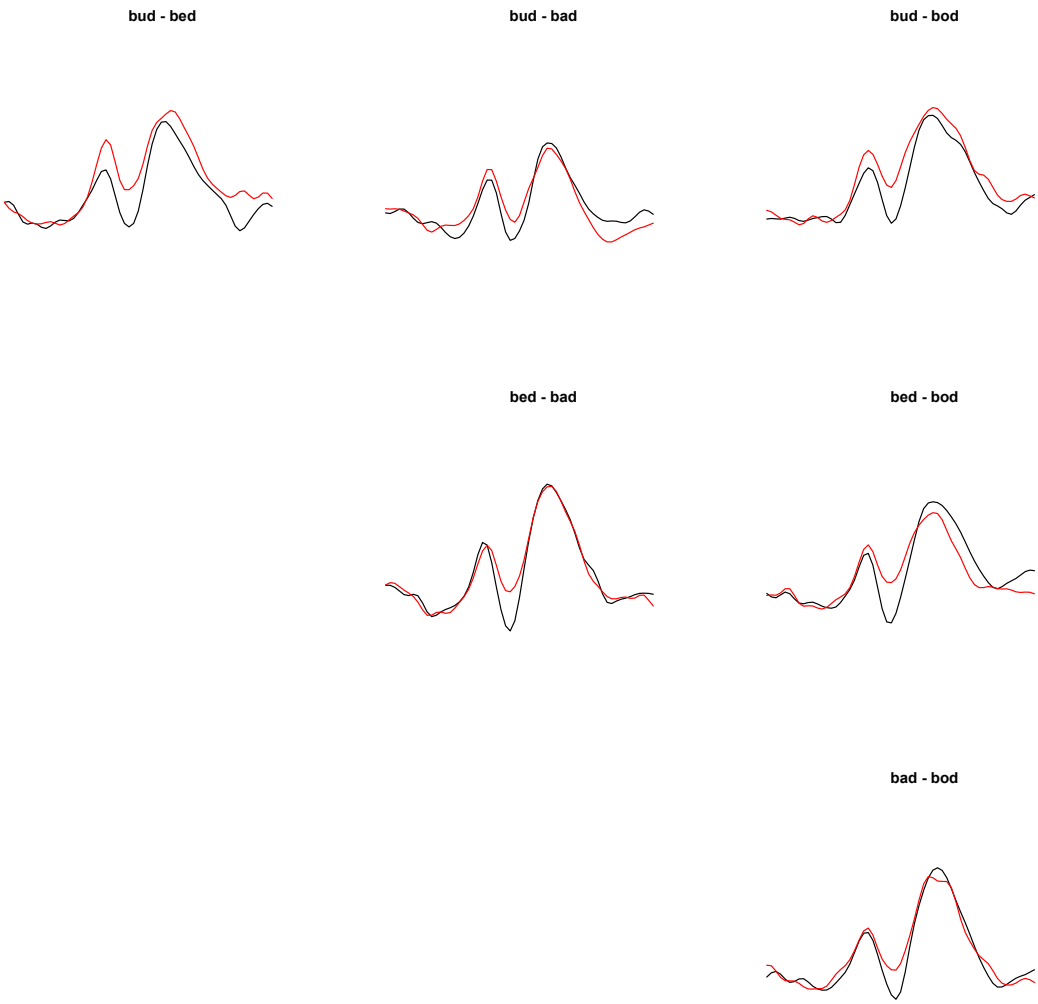

Figure 34: English vowels for L1 and L2 speakers at FCz (black=English L1 speakers, red=German L1 speakers)

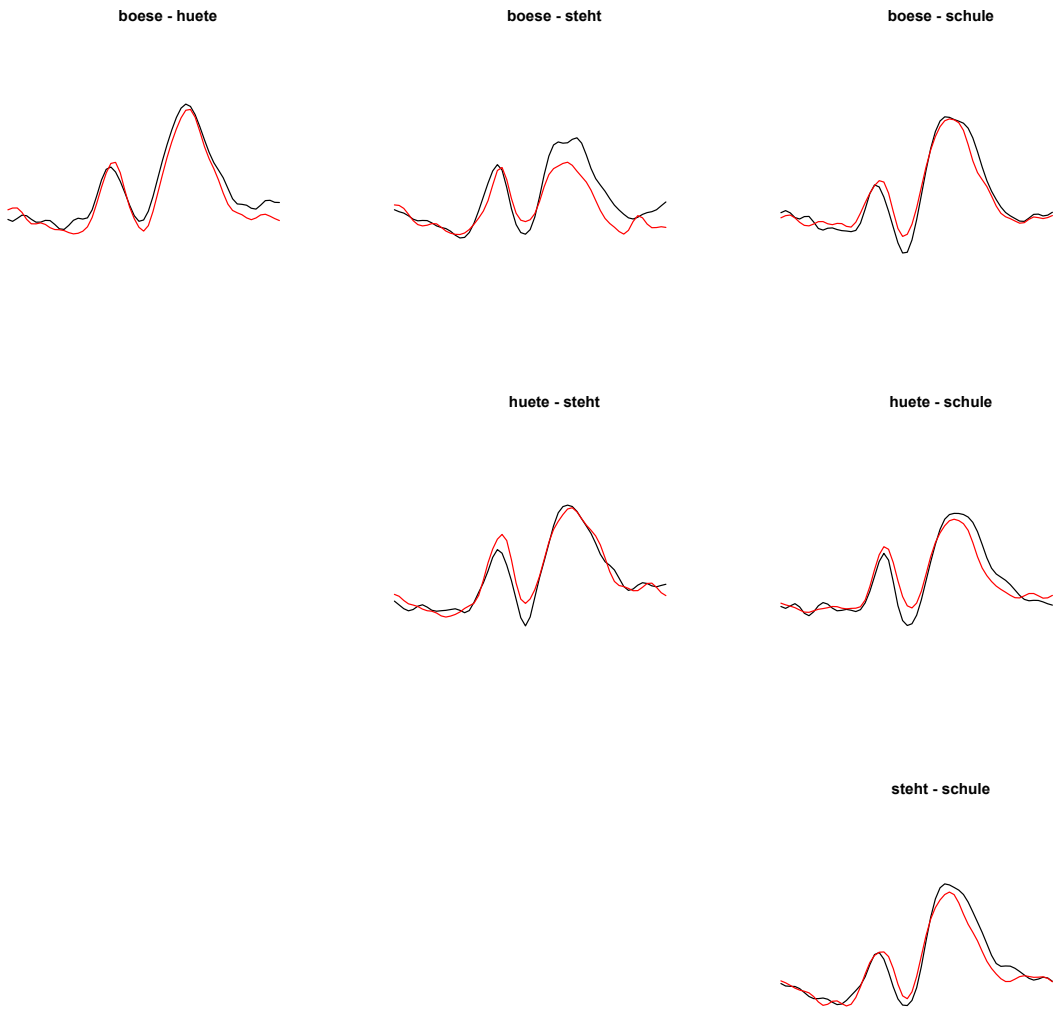

Figure 35: German vowels for L1 and L2 speakers at FCz (black=English L1 speakers, red=German L1 speakers)

2) FCz mixed models

One of the points of this second study was to be able to examine the cross-language difference more simply, so we only used an analysis with subject language and stimulus language as fixed factors, and subject and pair as random factors. Also, since we obtained cleaner waveforms due to a substantially larger number of trials/condition, we were able to assess P1, N1, and P2 separately,

in both magnitude and latency.

P1 magnitude at FCz

The first set of mixed-model analyses focused on the P1 response magnitude at FCz. There were no significant effects of subject or stimulus language on the P1 magnitude at electrode location FCz.

N1 magnitude at FCz

The next analysis investigated N1 response magnitude at FCz. There was an interaction of subject language and stimulus language ($F(1,388)=9.1$, $p<0.05$) with the N1 magnitude at FCz, in that the German L1 speakers had differing N1 magnitude responses to both the English and the German stimuli than the English L1 speakers – their amplitudes were higher at FCz (see Figure 37).

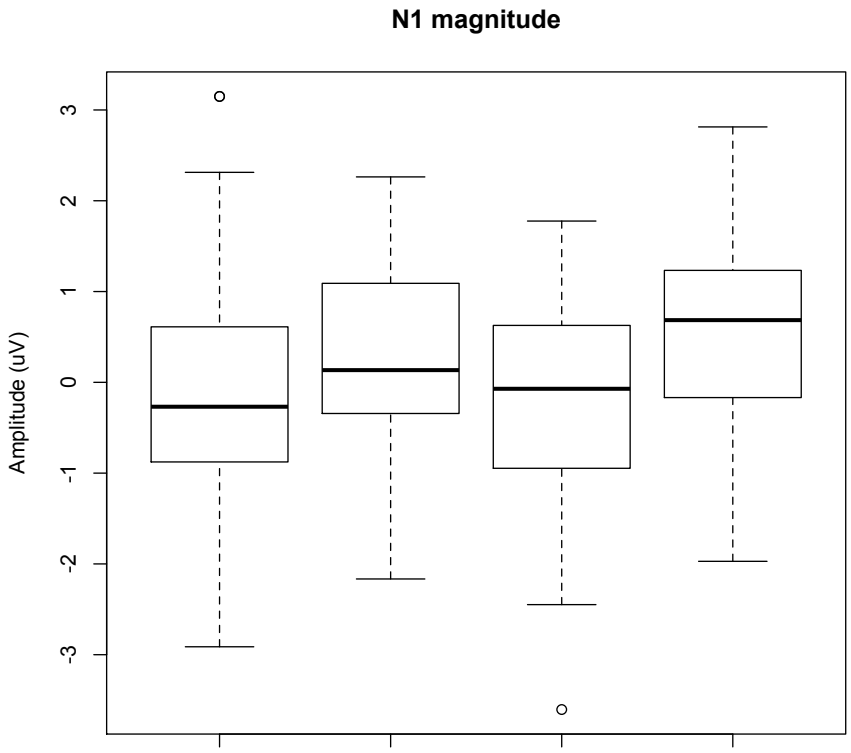

Figure 36: N1 magnitude at FCz

P2 magnitude at FCz

The next analysis focused on P2 response magnitude at FCz. There was an effect of pair language on the P2 magnitude at FCz ($F(1,388)=4.19$, $p<0.05$).

P1 latency at FCz

The next analysis investigated P1 response latency at FCz. None of the tested conditions (pair language and subject language) had an effect upon P1 latency at FCz.

N1 latency at FCz

There was an effect of pair language on the N1 latency at FCz ($F(1,388)=11.4$, $p<0.05$).

P2 latency at FCz

The following analysis focused on P2 latency at FCz. There was a significant effect of pair language on P2 latency at FCz ($F(1,388)=32.8$, $p<0.05$). Additionally, there was an interaction of pair language and subject language ($F(1,388)=6.3$, $p<0.05$) in that both the English and the German speakers had longer (but differing) latencies when responding to German stimuli, the English L1 speakers taking longer than the German L1 speakers (see Figure 38).

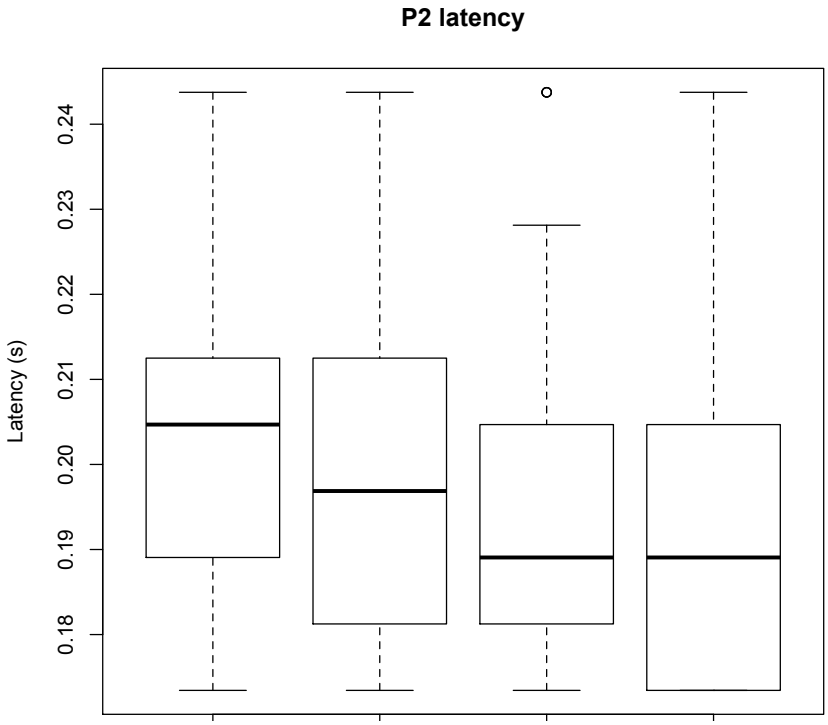

Figure 37: P2 latency at FCz

3) Analysis at P9

Below is the response matrix at P9 (which was the largest response of the temporoparietal cluster, as established through the previous ACC experiment) for the English vowel pairs (see Figure 39) and the German vowel pairs (see Figure 40), with black for English listeners and red for German listeners, positive-going waves plotted upwards. When looking at the English vowel pair matrix, it is immediately visible that certain vowel pairs show more of a cross-language difference between the two groups of speakers than others, specifically "bud-bed", "bud-bod", "bad-bod", "bed-bad". In

fact, the ACC responses for these stimulus pairs do not look like typical ACC responses at all, with a clear P1 and P2 being absent. These stimuli look more like the T-complex component described previously by Näätänen and Picton (1987) and Tonnquist-Uhlen (2003). We can see the early negative wave Na, followed by the positive wave Ta and the larger negative wave Tb in the L1 speaker response data, especially for vowel pairs "bed-bod" and "bad-bod". The L2 speaker data does not that much like a clear T-complex response, with the Na component being present for all vowel pairs, but the Ta component not always being clearly elicited. The Tb seems to be largely absent for the L2 speakers (particularly visible for the "bed-bod" vowel pair).

When looking at the German vowel pair matrix, it is immediately obvious that the T-complex components are much more pronounced than for the English vowel pairs, for both the English and the German L1 speakers. Additionally, both language groups seem to be more similar to each other. The main cross-language difference seems to lie in the Na component, with the German L1 speakers showing a larger Na magnitude than the English L1 speakers (particularly visible for the vowel pairs "huete-schule" and "boese-huete").

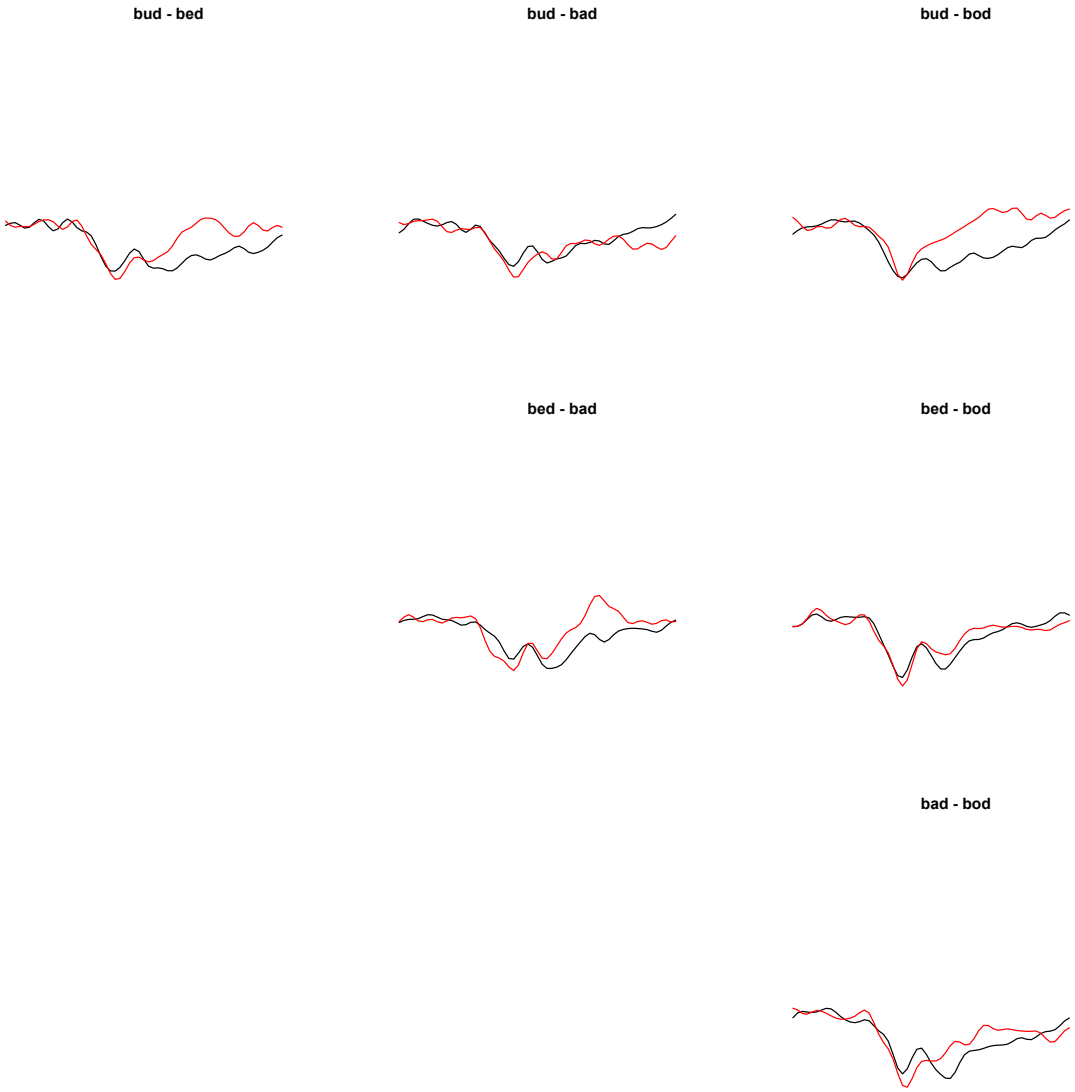

Figure 38: English vowels for L1 and L2 speakers at P9 (black=English L1 speakers, red= German L1 speakers)

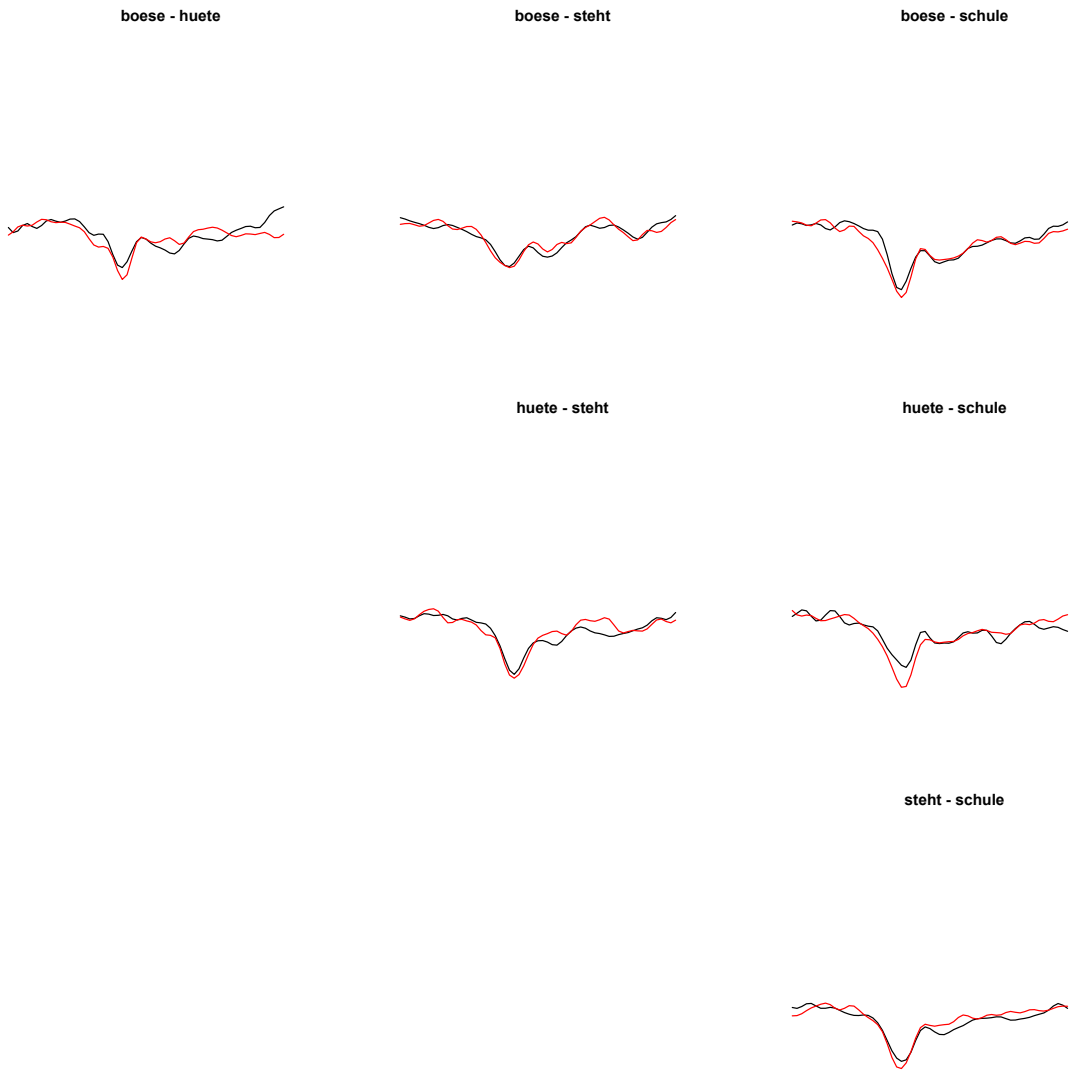

Figure 39: German vowels for L1 and L2 speakers at P9 (black=English L1 speakers, red=German L1 speakers)

4) P9 mixed models

N1 magnitude at P9

The first analysis investigated N1 response magnitude at P9. There was no effect of any of the tested conditions (pair language and subject language) on the N1 magnitude at electrode location P9.

N1 latency at P9

The next analysis focused on the N1 response latency at P9. There was an effect of pair language ($F(1,388)=24.8$, $p<0.05$) and an interaction between pair language and subject language ($F(1,388)=5.7$, $p<0.05$) on N1 latency at P9, in that English L1 speakers had a substantially longer N1 latency for English stimuli compared to German stimuli and also compared to the German speakers for both English and German stimuli (see Figure 41).

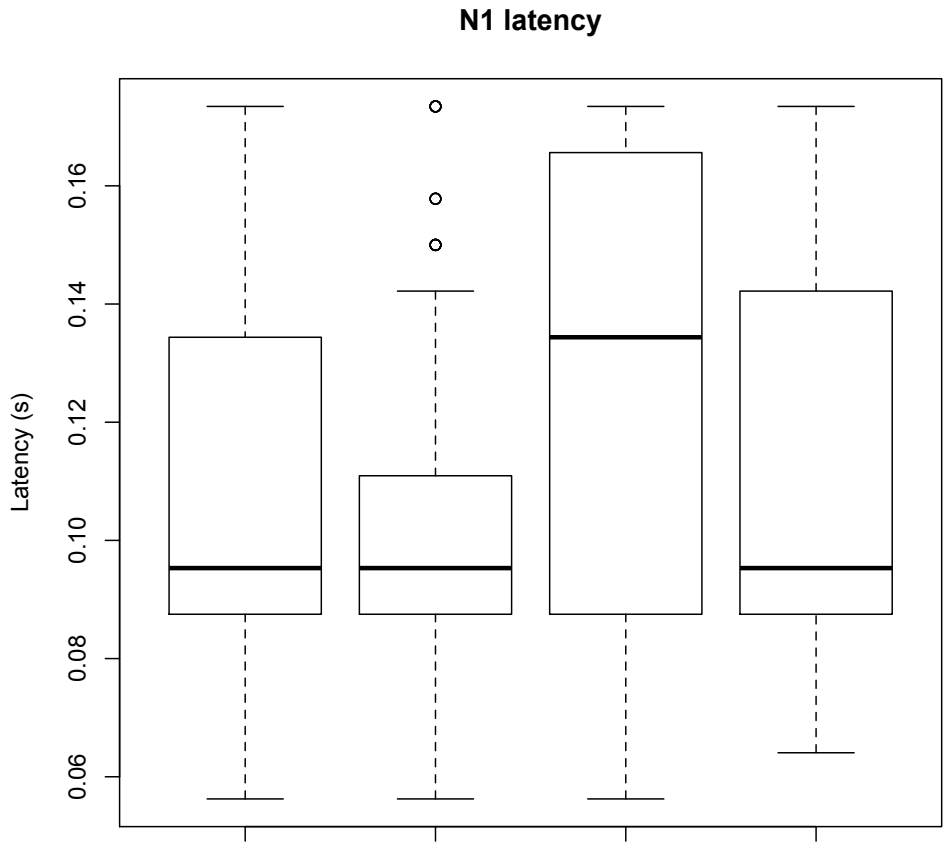

Figure 40: N1 latency at P9. Subject L1 English or German. Stimulus language English or German.

Discussion

The main finding of this set of EEG experiments was a cross-language difference for the ACC. Additionally, using a data-driven approach, we re-created the notion of a T-complex versus the standard ACC response. The ACC was not clearly visible in the temporoparietal response in the first EEG experiment. However, our novel data-driven approach was supported by a T-complex-like response visible in the temporoparietal region in the second EEG experiment. This T-complex-like response was strongest at P9, not at T7 and T8 as previous literature has found (Näätänen & Picton, 1987; Tonnquist-Uhlen et al., 2003), so it was not at the traditional T-complex location. At the traditional electrodes T7/T8 the responses were too small, which is why we used P9 for analysis instead.

Additionally to the T-complex, we also obtained a cross-language N1 effect when other researchers (Brint, 2012; Elangovan & Stuart, 2011; Martin et al., 2012) do not. This could be because a standard ACC design only uses 1-2 pairs and generally leads to habituation and flattening of the N1 response (Haenschel et al., 2005). The large amount of change between vowel pairs which was built into our experimental design was ideal for eliciting the N1. We had in fact designed our experiments from the outset to be suitable for N1 elicitation.

We achieved a cross-language N1 effect, but none for P2. This could be because both N1 and P2 are thought to rely on acoustic differences (Crowley & Colrain, 2004). N1 is thought to also rely on attention and P2 may rely on categorisation (Crowley & Colrain, 2004). If P2 really is driven by categorisation, then we should have been able to see cross-language differences in P2, but we did not. This finding is supported by studies which report only very small correlations between N1 and P2 (Crowley & Colrain, 2004).

The N1 relates to acoustic differences and is thought to be modulated by attention (Escera, Alho, Schroeger, & Winkler, 2000; Martin et al., 2012; Näätänen, 1975). So is the cross-language difference we are finding modulated by perceived acoustic difference or is it by modulating attention? The experimental setting is identical for both language groups, but the Germans are in a constant L2 setting in the UK, so maybe their general attention is modulated globally by being in a non-native environment. This could be due to their being used to expecting to hear and be exposed to L2 speech all the time (despite being spoken to and instructed exclusively in German during the experiments), resulting in a baseline change in N1 due to heightened global attention.

We tried a promising MDS design but did not get any results. However, ongoing work in our group (Iverson et al., 2013) shows that the MDS approach does, in principle, work with the ACC

paradigm. More trials and more vowel pairs would be necessary to get enough repetitions and a large enough vowel space to avoid over-fitting of the data. This could be achieved by having double the amount of EEG recording sessions.

We tested both Northern/Standard German and Southern German/Bairisch regional speakers. One might ask whether the regional accent differences between North and South could have had an effect on the behavioural and brain measures. This is a possibility. Due to an insufficient number of Southern German participants I could unfortunately not test for an effect of regional accent. However, I am bilingual in German and English and I spoke with each German participant in Standard German exclusively to establish whether they spoke Standard German (instead of only a Southern German dialect). Each native German participant spoke Standard German, the Southern German participants spoke with a slight regional accent.

Indeed, the effect of German regional accent may not be as large as one might think. Germany is a multidialectal society in which all younger generation native German speakers speak Standard German, orthography is in Standard German and most of the media and all education is in Standard German. Speakers are used to hearing and speaking Standard German in everyday life.

Standard German has the front rounded vowels /ʏ - yː - øː - œ/,

while in the main Southern German accents the front rounded vowels have been replaced by their corresponding front unrounded vowels /iː - ɪ - eː - ɛ/ (Barbour & Stevenson, 1990). Diphthongisation of monophthongs occurs and short vowels are transformed into long vowels and vice versa in Southern German dialects compared to Standard German (Koenig, 2007). However, all my participants were bidialectal in that they spoke Standard German as well as Southern German natively. As such, both phoneme systems were in their native phoneme inventory. As our experiment tested the differences between English and German vowel inventories, this should not have had a large effect on the results. Indeed, our results indicate that all native German participants showed robust ACC effects to English vowels.

Chapter 5: General discussion

We aimed to investigate several aspects of L2 auditory vowel processing in this thesis. Firstly, we were interested in at how low a level in automatic auditory processing cross-language differences are visible by investigating first the MMNm and then the ACC. We were interested in exploring how low down in neuronal processing we could go hierarchically and still get a cross-language difference in neuronal responses and neural network architecture. So far, the consensus has been that cross-linguistic differences only arise from the auditory cortex upwards. A study by Elangovan et al. (Elangovan & Stuart, 2011), using the P1-N1-P2 onset response to investigate voice onset time of consonants, found no cross-lingustic differences between L1 and L2 speakers, despite these subjects exhibiting differences in their phonetic category boundaries, depending on language background. The researchers inferred from these results that the onset P1-N1-P2 response is sensitive only to acoustic differences, not phonetic categorisation effects between languages. Additionally to cross-linguistic influences on relatively late pre-attentive processing with the MMN we found cross-linguistic influences on very early pre-attentive automatic auditory processing with the ACC. Our data are supported by the fact that very low-level auditory brain-stem processing shows effects of cross-linguistic

influence (Krishnan & Gandour, 2009), showing evidence of very early cross-linguistic influence on auditory pitch processing.

Secondly, we wanted to know what the effects of native language on neuronal architecture are. The MMN stimuli elicited different MMFs for each deviant type and showed a left-hemispheric dominance typical of speech stimuli. A study by Zevin et al. (Zevin et al., 2010) found L1 versus L2 lateralisation effects with the MMN, in that L1 speakers showed left-lateralisation and L2 speakers showed right-lateralisation effects. Our MMN results did not show these kinds of L1 vs L2 lateralisation effects between groups. Instead, activation was left-lateralised for all subjects, due to the stimuli being speech sounds (and not non-speech sounds). Our ACC data did not show any lateralisation effects. However, it is possible that it may have, if analysed using DCM.

Thirdly, we explored whether cross-language differences are visible in the MMNm. These kinds of cross-language effects have in fact already been found in the past for the MMN (Peltola et al., 2003), but we have shown cross-language effects by modeling with neuronal sources, which has not been done before. DCM fit a four dipole model to our L1 and L2 speaker MEG mismatch data. Each of our three deviants showed a different neural network compared to the standard. Increasing deviance lead to the increasing involvement of self connections and lateral connections. With the

deviants 2 and 3, the L2 speakers had greater self connections for the right superior temporal gyrus, this is an indication that the right hemisphere is supporting the left, possibly by focusing on frequency characteristics. Behaviourally, discriminating between phonemes was linked to D2, indicating a need for additional processing for frequency in the right hemisphere.

Fourthly, we looked at whether cross-language differences are visible in the ACC. The ACC response data clustered into two different data-driven scalp areas: Frontocentral and temporoparietal areas which are roughly comparable with the T-complex. The source for the T-complex is believed to be in the superior temporal gyrus. The first ACC experiment's responses for frontocentral N1 latency and magnitude indicated that categorisation effects are not related to the language-background status of a subject, but rather an individual's identification abilities in English. As these identification abilities can change and improve over time, this would imply that changes in the underlying neural network are gradual, take place over time and can possibly reach native-like status at the phoneme perception level. Temporoparietal N1 magnitude and latency were not affected by acoustic distance between vowels, but by individual categorisation abilities, indicating that it could be driven by phoneme differences. The T-complex is assumed to reflect features of linguistic processing (Chang et al., 2010; Friedrich, Schild, &

Roder, 2009; Schafer, Schwartz, & Martin, 2011). It shows cross-linguistic differences and is more negative for L1 speakers for onset P1-N1-P2 (Wagner et al., 2013). The T-complex for onset P1-N2-P2 is sensitive to acoustic, but not to phonological aspects of speech (Elangovan & Stuart, 2011; Sharma & Dorman, 2000; Wagner et al., 2013). However, our data (using the ACC) indicates that acoustic difference has no effect on either temporoparietal N1 magnitude or latency. In fact, our data indicates that tempoparietal N1 is not affected by acoustic distance between vowels but rather by individual categorisation abilities which could mean that it is therefore being driven by phoneme differences.

Fifthly, we investigated what behavioural measures tell us about neuronal processing and neuronal architecture. The perception and production tasks showed cross-language effects and, surprisingly, no relationship between production and perception abilities. The link between category discrimination and MMF for our experimental deviant D2 could indicate the phoneme status of a speech sound in a listener's brain.

For future research, it would be informative to perform dynamic causal modeling on ACC data in order to find out whether there are further similarities in the underlying neuronal network of both MMN and ACC and to explore the links between ACC DCM and the behavioural data collected. DCM could be explored by running

another ACC experiment, this time with fewer vowel pairs for feasibility reasons. Preferably 2-3 vowel pairs as each vowel pair very considerably adds to the already very lengthy processing time for DCM.

Looking at practical issues for future research, it is worth mentioning that the ACC elicits a response with a larger amplitude than MMN (Martin & Boothroyd, 1999). This leads to improved response detectability on an individual subject basis. This is especially useful in the context of L2 research where individual subjects differ from person to person in their language abilities and proficiency. Therefore, it would be more possible to explore individual differences between subjects using the ACC as compared to the MMN, which is better for looking at group differences. The ACC is also more efficient for experimental use than MMN, as every stimulus change counts as a trial. With the MMN, at least two standards are necessary before a deviant, every deviant counting as a trial (Martin, Boothroyd, Ali, & Leach-Berth, 2010).

In this novel set of brain-imaging experiments, which explored L2 auditory vowel processing using different brain-imaging techniques (MEG and EEG) and DCM modeling, we were pushing the boundaries of how low and how automatic one can get cross-language differences in auditory vowel processing. The MMN is a relatively basic automatic brain response, but it still requires a

design which builds expectations and predictions in the brain. The ACC response, however, is only based on sound changes and does not involve any expectations or predictions. The ACC response happens too rapidly for subjects to be able to think anything apart from that they are hearing many different sounds. And yet, we found cross-language differences not only for the MMNm, using both a conventional MMF peak analysis and dynamic causal modelling, but also for the ACC. This implies that the native language has an effect at a very low speech processing level.

The difference between low and high ability speakers was visible in neural architecture, in that low ability speakers relied more on the right hemisphere to process the D2 mismatch response and showed less efficient L2 speech sound processing. L2 speakers who had been resident in the UK longest showed an L1-like neural architecture. The fact that highly proficient L2 speakers are indistinguishable from L1 speakers at the neural architecture level for D2 seems to indicate that the human brain is highly adaptable and retains perceptual plasticity into adulthood by learning to categorise and identify a novel sound and integrate it into an automatically processed, native-like L2 phoneme inventory.

Our DCM analysis showed that highly proficient L2 speakers were indistinguishable from L1 speakers with respect to their brain connectivity architecture. This could be due to two possibilities:

Either certain people possess brain connectivity architecture which predisposes them to high ability L2 vowel perception or it is possible to achieve native-like L2 vowel perception. The correlation of D2 with RSTG and our behavioural measure of exposure in the UK supports the latter option; the cross-language differences seem to arise from exposure and training, not aptitude. This suggests that increasing exposure to the L2 in an immersive environment can lead to native-like processing of L2 vowels over time.

Literature

Aguilar Ayala, Y., & Malmierca, M.S. (2013). Stimulus-specific adaptation and deviance detection in the inferior colliculus. *Frontiers in Neural Circuits, 6*, 1662-5110.

Alink, A., Schwiedrzik, C.M., Kohler, A., Singer, W., & Muckli, L. (2010). Stimulus Predictability Reduces Responses in Primary Visual Cortex. *The Journal of Neuroscience, 30*(8), 2960-2966.

Barbour, Stephen, & Stevenson, Patrick. (1990). *Variation in German: A Critical Approach to German Sociolinguistics*: Cambridge University Press.

Bidelman, G.M., Gandour, J.T., & Krishnan, A. (2011). Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *Journal of Cognitive Neuroscience, 23*(2), 425-434.

Bidelman, Gavin M., Gandour, J.T., & Krishnan, A. (2011). Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. *Brain and Cognition, 77*(1), 1-10.

Bidelman, Gavin M., Moreno, Sylvain, & Alain, Claude. (2013). Tracing the emergence of categorical speech perception in the human auditory system. *NeuroImage, 79*(0), 201-212.

Birdsong, D. (1999). *Second Language Acquisition and the Critical Period Hypothesis*. Mahwah, NJ: Lawrence Erlbaum.

Bishop, D.V., Brown, B.B., & Robson, J. (1990). The relationship between phoneme discrimination, speech production, and language comprehension in cerebral-palsied individuals. *Journal of Speech and Hearing Research, 33*, 210-219.

Bradlow, A., Akahane-Yamada, R., Pisoni, D., & Tohkura, Y. . (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics, 61*, 977-985.

Bradlow, A., Pisoni, D., Akabane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America, 101*, 2299-2310.

Brint, E. (2012). *Cross Language Differences in Cortical Auditory Evoked Responses.* (MRes), University College London, London.

Brown, C.J., Etler, C., He, S., O'Brien, S., Erenberg, S., Kim, J.R., . . . Abbas, P.J. (2008). The electrically evoked auditory change complex: preliminary results from nucleus cochlear implant users. *Ear and Hearing, 29*(5), 704-717.

Buechel, C., & Friston, K.J. (1997). Modulation of connectivity in visual pathways by attention: Cortical interactions evaluated

with structural equation modelling and fMRI. *Cerebral Cortex, 7*, 768-778.

Chang, E., Rieger, J., Johnson, K., Berger, M., Barbaro, N., & Knight, R.T. (2010). Categorical speech representation in human superior temporal gyrus. *Nature Neuroscience, 13*, 1428-1432.

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., & Näätänen, R. (1998). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience, 1*(5), 351-353.

Cowan, N., Winkler, I., Teder, W., & Näätänen, R. (1993). Memory prerequisites of the mismatch negativity in the auditory event-related potential. *Journal of Experimental Psychology: Language, Memory and Cognition, 19*, 909-921.

Crowley, K.E., & Colrain, I.M. (2004). A review of the evidence for P2 being an independent component process: age, sleep and modality. *Clinical Neurophysiology, 115*(4), 732-744.

Cutler, A., Weber, A., Smits, R., & Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *Journal of the Acoustical Society of America, 116*(6), 3668-3678.

David, O., Kiebel, S.J., Harrison, L.M., Mattout, J., Kilner, J.M., & Friston, K.J. (2006). Dynamic causal modeling of evoked

responses in EEG and MEG. *NeuroImage, 30*(4), 1255-1272.

Davis, P.A. (1939). Effects of acoustic stimuli on the waking human brain. *Neurophysiology, 2*(494-499).

Dehaene-Lambertz, G. (1997). Electrophysiological correlates of categorical phoneme perception in adults. *NeuroReport, 8*, 919-924.

Dehaene-Lambertz, G., Dupoux, E., & Gout, A. (2000). Electrophysiological Correlates of Phonological Processing: A Cross-linguistic Study. *Journal of Cognitive Neuroscience, 12*(4), 635-647.

DeKeyser, R. (1997). Beyond explicit rule learning. *Studies in Second Language Acquisition, 19*, 195-221.

DeKeyser, R. (2001). Automaticity and automatization. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 125-151). New York: Cambridge University Press.

den Ouden, H.E.M., Daunizeau, J., Roiser, J., Friston, K.J., & Stephan, K.E. (2010). Striatal Prediction Error Modulates Cortical Coupling. *The Journal of Neuroscience, 30*(9), 3210-3219.

Diaz, B., Baus, C., Escera, C., Costa, A., & Sebastien-Galles, N. (2008). Brain potentials to native phoneme discrimination reveal the origin of individual differences in learning the sounds of a second language. *Proceedings of the National*

Academy of Sciences, 105(42), 16083-16088.

Eimas, P.D., Siqueland, E.R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science, 171*(968), 303-306.

Elangovan, S., & Stuart, A. (2011). A cross-linguistic examination of cortical auditory evoked potentials for a categorical voicing contrast. *Neuroscience Letters, 490*(2), 140-144.

Escera, C., Alho, K., Schroeger, E., & Winkler, I. (2000). Involuntary attention and distractability as evaluated with event-related brain potentials. *Audiology & Neuro-Otology, 5*, 151-166.

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *The European Journal of Neuroscience, 15*(2), 399-402.

Favreau, M., & Segalowitz, N. (1983). Automatic and controlled processes in the first- and second-language reading of fluent bilinguals. *Memory and Cognition, 11*(6), 565-574.

Felleman, D.J., & Van Essen, D.C. (1991). Distributed Hierarchical Processing in the Primate Cerebral Cortex. *Cerebral Cortex, 1*(1), 1-47.

Flege, J.E. (1999). *The relation between L2 production and perception.* Paper presented at the XIVth International Congress of Phonetics Sciences, Berkeley, CA: Department of

Linguistics, Univ. of California at Berkeley.

Friedrich, C.K., Schild, U., & Roder, B. (2009). Electrophysiological indices of word fragment priming allow characterizing neural stages of speech recognition. *Biological Psychology, 80*, 105-113.

Friesen, L.M., & Tremblay, K.L. (2006). Acoustic Change Complexes Recorded in Adult Cochlear Implant Listeners. *Ear and Hearing, 27*(6), 678-685.

Friston, K.J. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences, 360*(1456), 815-836.

Friston, K.J. (2009). Causal Modelling and Brain Connectivity in Functional Magnetic Resonance Imaging. *PLoS Biol, 7*(2), e1000033.

Friston, K.J., Harrison, L., & Penny, W. (2003). Dynamic causal modeling. *NeuroImage, 19*(4), 1273-1302.

Friston, K.J., & Kiebel, S.J. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1521), 1211-1221.

Friston, K.J., & Stephan, K.E. (2007). Free-Energy and the Brain. *Synthese, 159*(3), 417-458.

Fu, Q.-J. , Galvin, J.J., Wang, X., & Nogaki, G. . (2005). Moderate auditory training can improve speech performance of adult

cochlear implant users. *Acoustics Research Letters Online, 6*, 106-111.

Galantucci, Bruno, Fowler, Carol, & Turvey, M.T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin Review, 13*(3), 361-377.

Gao, S., Hu, J., Gong, D., Chen, S., Kendrick, K.M., & Yao, D. (2012). Integration of Consonant and Pitch Processing as Revealed by the Absence of Additivity in Mismatch Negativity. *PLoS ONE, 7*(5), e38289.

Garrido, M.I., Friston, K.J., Kiebel, S.J., Stephan, K.E., Baldeweg, T., & Kilner, J.M. (2008). The functional anatomy of the MMN: A DCM study of the roving paradigm. *NeuroImage, 42*(2), 936-944.

Garrido, M.I., Kilner, J.M., Kiebel, S.J., & Friston, K.J. (2007). Evoked brain responses are generated by feedback loops. *Proceedings of the National Academy of Sciences, 104*(52), 20961-20966.

Garrido, M.I., Kilner, J.M., Stephan, K.E., & Friston, K.J. (2009). The mismatch negativity: A review of underlying mechanisms. *Clinical Neurophysiology, 120*(3), 453-463.

Glasberg, Brian R., & Moore, Brian C. J. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research, 47*(1–2), 103-138.

Gockel, H., Carlyon, R., Mehta, A., & Plack, C. (2011). The Frequency Following Response (FFR) May Reflect Pitch-Bearing Information But is Not a Direct Representation of Pitch. *Journal of the Association for Research in Otolaryngology, 12*, 767-782.

Golestani, N., Molko, N., Dehaene, S., LeBihan, D., & Pallier, C. (2007). Brain structure predicts the learning of foreign speech sounds. *Cerebral Cortex, 17*, 575-582.

Golestani, N., Paus, T., & Zatorre, R.J. (2002). Anatomical correlates of learning novel speech sounds. *Neuron, 35*, 997-1010.

Golestani, N., & Zatorre, R.J. (2009). Individual differences in the acquisition of second language phonology. *Brain and Language, 109*, 55-67.

Grimm, S., & Escera, C. (2012). Auditory deviance detection revisited: Evidence for a hierarchical novelty system. *International Journal of Psychophysiology, 85*(1), 88-92.

Haenschel, C., Vernon, D., Dwivedi, P., Gruzelier, J., & Baldeweg, T. (2005). Event-related brain potential correlates of human auditory sensory memory-trace formation. *Journal of Neuroscience, 25*(9), 10494-10501.

Hari, R., Pelizzonne, M., Makela, J.P., Hallstrom, J., Leinonen, L, & Lounasmaa, O.V. (1987). Neuromagnetic responses of the

human auditory cortex to on- and offsets of noise bursts. *International Journal of Audiology, 26*(1), 31-43.

Hattori, K. (2009). *Perception and Production of English /r/-/l/ by Adult Japanese Speakers.* (PhD), University College London, London.

Hattori, K., & Iverson, P. (2009). English /r/-/l/ category assimilation by Japanese adults: individual differences and the link to identification accuracy. *Journal of the Acoustical Society of America, 125*(1), 469-479.

Hazan, V., Sennema, A., Iba, M., & Faulkner, A. . (2005). Effect of audiovisual perceptual training on the perception and production of consonants in Japanese learners of English. *Speech Communication, 47*, 360-378.

Heeren, W.F.L., & Schouten, M.E.H. (2008). Perceptual development of phoneme contrasts: How sensitivity changes along acoustic dimensions that contrast phoneme categories. *Journal of the Acoustical Society of America, 124*, 2291-2302.

Hickok, G. (2013). The architecture of speech production and the role of the phoneme in speech processing. *Language, Cognition and Neuroscience, 29*(1), 2-20.

Hickok, G., Costanzo, M., Capasso, R., & Miceli, G. (2011). The role of Broca's area in speech perception: Evidence from aphasia revisited. *Brain and Language, 119*(3), 214-220.

Hickok, G., Okada, K., Barr, W., Pa, J., Rogalsky, C., Donnelly, K., . . . Grant, A. (2008). Bilateral capacity for speech sound processing in auditory comprehension: Evidence from Wada procedures. *Brain and Language, 107*(3), 179-184.

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition, 92*(1-2), 67-99.

Hickok, Gregory. (2012). The cortical organization of speech processing: Feedback control and predictive coding the context of a dual-stream model. *Journal of Communication Disorders, 45*(6), 393-402.

Hickok, Gregory, & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience, 8*, 393-402.

Hillyard, S. A., & Kutas, M. (1983). Electrophysiology of Cognitive Processing. *Annual Review of Psychology, 34*(1), 33-61.

Huang, Y., & Rao, R.P.N. (2011). Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science, 2*(5), 580-593.

Iverson, P., Bernstein, L.E., & Auer, E.T. (1998). Modeling the interaction of phonemic intelligibility and lexical structure in audiovisual word recognition. *Speech Communication, 26*(1-2), 45-63.

Iverson, P., & Evans, B. (2007). Learning English vowels with different first-language vowel systems: Perception of formant

targets, formant movement, and duration. *Journal of the Acoustical Society of America, 122*(5), 2842-2854.

Iverson, P., & Evans, B. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *Journal of the Acoustical Society of America, 126*, 866-877.

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *The Journal of the Acoustical Society of America, 118*(5), 3267-3278.

Iverson, P., Kuhl, P.K. , Akahane-Yamada, R., Diesch, E., Tohkura, Y., & Kettermann, A. (2003). A Perceptual Interference Account of Acquisition Difficulties for Non-native Phonemes. . *Cognition, 87*, B47–B57.

Iverson, P., Mulyak, M., & Wagner, A. (2013). Multidimensional scaling of English fricatives using the acoustic change complex of electroencephalogram recordings. *Journal of the Acoustical Society of America, 133*(5), 3605.

Iverson, P., Pinet, M., & Evans, B.G. (2011). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics, FirstView*, 1-16.

Jääskeläinen, I.P., Ahveninen, J., Bonmassar, G., Dale, A.M.,

Ilmoniemi, R.J., Levänen, S., . . . Belliveau, J.W. (2004). Human posterior auditory cortex gates novel sounds to consciousness. *Proceedings of the National Academy of Sciences of the United States of America, 101*(17), 6809-6814.

Jacobsen, T., Schröger, E., & Alter, K. (2004). Pre-attentive perception of vowel phonemes from variable speech stimuli. *Psychophysiology, 41*(4), 654-659.

Jansen, B.H., & Rit, V.G. (1995). Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. *Biological Cybernetics, 73*(4), 357-366.

Jaramillo, M., Ilvonen, T., Kujala, T., Alku, P., Tervaniemi, M., & Alho, K. (2001). Are different kinds of acoustic features processed differently for speech and non-speech sounds? *Cognitive Brain Research, 12*, 459–466.

Javitt, D.C., Steinschneider, M., Schroeder, C.E., & Arezzo, J.C. (1996). Role of cortical N-methyl-D-aspartate receptors in auditory sensory memory and mismatch negativity generation: implications for schizophrenia. *Proceedings of the National Academy of Sciences, 93*(21), 11962-11967.

Jemel, B., Achenbach, C., Müller, B.W., Röpcke, B., & Oades, R.D. (2002). Mismatch Negativity Results from Bilateral

Asymmetric Dipole Sources in the Frontal and Temporal Lobes. *Brain Topography, 15*(1), 13-27.

Jewett, Don, & Williston, John. (1971). Auditory-evoked Farfields averaged from the scalp of humans *Brain, 94*, 681-696.

Kaan, E. (2007). Event-Related Potentials and Language Processing: A Brief Overview. *Language and Linguistics Compass, 1*(6), 571-591.

Kasai, K., Yamada, H., Kamio, S., Nakagome, K., Iwanami, A., Fukuda, M., . . . Ueno, S. (2001). Brain lateralization for mismatch response to across- and within-category change of vowels. *NeuroReport, 12*(11), 2467-2471.

Kiebel, S.J., Daunizeau, J., Phillips, C., & Friston, K.J. (2008). Variational Bayesian inversion of the equivalent current dipole model in EEG/MEG. *NeuroImage, 39*(2), 728-741.

Kiebel, S.J., Garrido, M.I., & Friston, K.J. (2007). Dynamic causal modelling of evoked responses: The role of intrinsic connections. *NeuroImage, 36*(2), 332-345.

Kiebel, S.J., Garrido, M.I., Moran, R., Chen, C.-C., & Friston, K.J. (2009). Dynamic causal modeling for EEG and MEG. *Human Brain Mapping, 30*(6), 1866-1876.

Kiebel, S.J., Garrido, M.I., Moran, R.J., & Friston, K.J. (2008). Dynamic causal modelling for EEG and MEG. *Cognitive Neurodynamics, 2*(2), 121-136.

Kim, J.-R., Brown, C.J., Abbas, P.J., Etler, C.P., & O'Brien, S. (2009). The Effect of Changes in Stimulus Level on Electrically Evoked Cortical Auditory Potentials. *Ear and Hearing, 30*(3), 320-329.

Klatt, D.H., & Klatt, L.C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America, 87*, 820-857.

Koenig, Werner. (2007). *dtv-Atlas: Deutsche Sprache* (17 ed.). Muenchen: Deutscher Taschenbuch Verlag.

Krishnan, A., & Gandour, J.T. (2009). The role of the auditory brainstem in processing linguistically-relevant pitch patterns. *Brain and Language, 110*(3), 135-148.

Krishnan, A., Swaminathan, J., & Gandour, J.T. . (2009). Experience dependent enhancement of linguistic pitch representation in the brainstem is not specific to a speech context. . *Journal of Cognitive Neuroscience, 21*(6), 1092-1105.

Krishnan, A., Xu, Y., Gandour, J., & Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research, 25*, 161-168.

Kuhl, P.K., & Iverson, P. (1995). Linguistic experience and the perceptual magnet effect. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-*

Language Research (pp. 121-154). York: Timonium.

Kuhl, P.K., & Miller, J.D. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science, 190*, 69-72.

Kuhl, P.K., Williams, K.A., Lacerda, F., & Lindblom, B. (1992). Language experience alters phonetic perception in infants by 6 months of age. *Science, 255*, 606-608.

Lambacher, S., Martens, W., Kakehi, K., Marasinghe, C., & Molholt, G. . (2005). The effects of identification training on the identification and production of American English vowels by

native speakers of Japanese. *Applied Psycholinguistics, 26*, 227-247.

Lane, Harlan. (1965). The motor theory of speech perception: A critical review. *Psychological Review, 72*(4), 275-309.

Leff, A.P., Iverson, P., Schofield, T.M., Kilner, J.M., Crinion, J.T., Friston, K.J., & Price, C.J. (2009). Vowel-specific mismatch responses in the anterior superior temporal gyrus: An fMRI study. *Cortex, 45*(4), 517-526.

Lehmann, Alexandre, & Schönwiesner, Marc. (2014). Selective Attention Modulates Human Auditory Brainstem Responses: Relative Contributions of Frequency and Spatial Cues *PLoS ONE, 9*(1), e85442.

Levanen, S., Ahonen, A., Hari, R., McEvoy, L., & Sams, M. (1996).

Deviant auditory stimuli activate human left and right auditory cortex differently. *Cerebral Cortex, 6*, 288-296.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. . (1967). Perception of the speech code. *Psychological Review, 74*(6), 431-461.

Liberman, Alvin, & Mattingly, Ignatius. (1985). The motor theory of speech perception revised. *Cognition, 21*(1), 1-36.

Liberman, Alvin, & Whalen, Doug. (2000). On the relation of speech to language. *Trends in Cognitive Sciences, 4*(5).

Liégeois-Chauvel, C., Musolino, A., Badier, J. M., Marquis, P., & Chauvel, P. (1994). Evoked potentials recorded from the auditory cortex in man: evaluation and topography of the middle latency components. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section, 92*(3), 204-214.

Litvak, V., Mattout, J., Kiebel, S.J., Phillips, C., Henson, R.N.A., Kilner, J., . . . Friston, K.J. (2011). EEG and MEG data analysis in SPM8. *Computational Intelligence and Neuroscience*, 852-961.

Logan, J.S., Lively, S.E., & Pisoni, D.B. (1991). Training Japanese listeners to identify English /r/ and /l/: a first report *Journal of the Acoustical Society of America, 89*, 874-886.

Luo, H, Ni, J T , Li, Z H , Li, X O , Zhang, D R , Zeng, F G, & Chen,

L. (2006). Opposite patterns of hemisphere dominance for early auditory processing of lexical tones and consonants. *PNAS, 103*(51), 19558-19563.

Martin, B.A., & Boothroyd, A. (1999). Cortical, auditory, event-related potentials in response to periodic and aperiodic stimuli with the same spectral envelope. *Ear and Hearing, 20*(1), 33-44.

Martin, B.A., & Boothroyd, A. (2000). Cortical, auditory, evoked potentials in response to changes of spectrum and amplitude. *Journal of the Acoustical Society of America, 107*(4), 2155-2161.

Martin, B.A., Boothroyd, A., Ali, D., & Leach-Berth, T. (2010). Stimulus Presentation Strategies for Eliciting the Acoustic Change Complex: Increasing Efficiency. *Ear and Hearing, 31*(3), 356-366.

Martin, B.A., Shafer, V., Wroblewski, M., & An, L.J. (2012). Functional significance of the acoustic change complex, mismatch negativity, and P3a for vowel processing in native-English listeners and late learners of English. *The Journal of the Acoustical Society of America, 132*(3), 1935-1935.

Martin, B.A., Tremblay, K.L., & Korczak, P. (2008). Speech Evoked Potentials: From the Laboratory to the Clinic. *Ear and Hearing, 29*(3), 285-313.

Martin, B.A., Tremblay, K.L., & Stapells, D.R. (2007). Principals and applications of cortical auditory evoked potentials. In R. F. Burkard, M. Don & J. J. Eggermont (Eds.), *Auditory evoked potentials: basic principles and clinical application* (pp. 482-507). Philadelphia: PA: Lippincott Williams and Wilkins.

McIntosh, A.R., & Gonzalez-Lima, F. (1994). Structural Equation Modeling and Its Application to Network Analysis in Functional Brain Imaging. *Human Brain Mapping, 2*, 2-22.

Mechelli, A., Price, C.J., Noppeney, U., & Friston, K.J. (2003). A dynamic causal modeling study on category effects: bottom-up or top-down mediation? *Journal of Cognitive Neuroscience, 15*(7), 925-934.

Mesgarani, N., & Chang, E.F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature, 485*, 233–236.

Moore, B.C., Glasberg, B.R., & Bear, T. (1997). A model for the prediction of thresholds, loudness, and partial loudness. *Journal of the Audio Engineering Society, 45*, 224-240.

Naatanen, R, Paavilainen, P, Rinne, T, & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clinical Neurophysiology, 118*, 2544–2590.

Näätänen, R. (1975). Selective attention and evoked potentials in

humans: A critical review. *Biological Psychology, 2*, 237-307.

Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology, 38*(1), 1-21.

Näätänen, R. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology, 118*, 2544-2590.

Näätänen, R., Gaillard, A.W.K., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica, 42*, 313-329.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., . . . Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature, 358*, 392-397.

Näätänen, R., Paavilainen, P., Tiitinen, H., Jiang, D., & Alho, K. (1993). Attention and mismatch negativity. *Psychophysiology, 30*, 436-450.

Näätänen, R., & Picton, T.W. (1987). The N1 Wave of the Human Electric and Magnetic Response to Sound: A Review and an Analysis of the Component Structure. *Psychophysiology, 24*(4), 375-425.

Näätänen, R., Schröger, E., Karakas, S., Tervaniemi, M., &

Paavilainen, P. (1993). Development of a memory trace for a complex sound in the human brain. *NeuroReport, 4*, 503-506.

Newell, A., & Rosenbloom, P. (1981). Mechanisms of Skill Acquisition and the Law of Practice. In J. Anderson (Ed.), *Cognitive Skills and their Acquisition* (pp. 1-55). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Nunez, P.L, & Srinivarsan, R. (2006). *Electric fields of the brain*. Oxford: Oxford University Press.

Okamoto, Hidehiko, & Kakigi, Ryusuke. (2013). Hemispheric Asymmetry of Auditory Mismatch Negativity Elicited by Spectral and Temporal Deviants: A Magnetoencephalographic Study. *Brain Topography*, 1-8. doi: 10.1007/s10548-013-0347-1

Oliver, G., Gullberg, M., Hellwig, F., Mitterer, H., & Indefrey, P. (2012). Acquiring L2 sentence comprehension: A longitudinal study of word monitoring in noise. *Bilingualism: Language and Cognition, 15*(4), 841-857.

Opitz, B., Mecklinger, A., Friederici, A.D., & von Cramon, D.Y. (1999). The Functional Neuroanatomy of Novelty Processing: Integrating ERP and fMRI Results. *Cerebral Cortex, 9*(4), 379-391.

Opitz, B., Rinne, T., Mecklinger, A., von Cramon, D.Y., & Schröger, E. (2002). Differential Contribution of Frontal and Temporal

Cortices to Auditory Change Detection: fMRI and ERP Results. *NeuroImage, 15*(1), 167-174.

Ostroff, J.M., Martin, B.A., & Boothroyd, A. (1998). Cortical Evoked Response To Acoustic Change Within A Syllable. *Ear and Hearing, 19*(4), 290-297.

Peltola, M.S., Kujala, T., Tuomainen, J., Ek, M., Aaltonen, O., & Näätänen, R. (2003). Native and foreign vowel discrimination as indexed by the mismatch negativity (MMN). *Neuroscience Letters, 1*(27), 25-28.

Peltola, M.S., Tuomainen, O., Koskinen, M., & Aaltonen, O. (2007). The Effect of Language Immersion Education on the Pre-Attentive Perception of Native and Non-Native Vowel Contrasts. *Journal of Psycholinguistic Research, 36*, 15-23.

Penny, W.D., Stephan, K.E., Daunizeau, J., Rosa, M.J., Friston, K.J., Schofield, T.M., & Leff, A.P. (2010). Comparing Families of Dynamic Causal Models. *PLoS Comput Biol, 6*(3), e1000709.

Penny, W.D., Stephan, K.E., Mechelli, A., & Friston, K.J. (2004). Comparing dynamic causal models. *NeuroImage, 22*(3), 1157-1172.

Picton, T.W., Hillyard, S.A., Krausz, H.I., & Galambos, R. (1974). Human auditory evoked potentials. I: Evaluation of components. *Electroencephalography and Clinical Neurophysiology, 36*(0), 179-190.

Pitt, M.A., & Myung, I.J. (2002). When a good fit can be bad. *Trends in Cognitive Sciences, 6*(10), 421-425.

Pulvermüller, F. (2003). Automatic processing of grammar in the human brain as revealed by the mismatch negativity. *NeuroImage, 20*, 159.

Pulvermüller, F., Kujala, T., Shtyrov, Y., Simola, J., Tiitinen, H., Alku, P., . . . Näätänen, R. (2001). Memory Traces for Words as Revealed by the Mismatch Negativity. *NeuroImage, 14*, 607–616.

Rao, R.P.N., & Ballard, D.H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci, 2*(1), 79-87.

Ringbom, H. (1992). On L1 Transfer in L2 Comprehension and L2 Production. *Language Learning, 42*(1), 85-112.

Rinne, T., Alho, K., Ilmoniemi, R. J., Virtanen, J., & Näätänen, Risto. (2000). Separate time behaviors of the temporal and frontal MMN sources. *NeuroImage, 12*, 14-19.

Rinne, T., Antila, S., & Winkler, I. . (2001). Mismatch negativity is unaffected by top-down predictive information. *NeuroReport, 12*(10), 2209-2213.

Rinne, Teamu, Alho, Kimmo, Alku, Paavo, Holi, Markus, Sinkkonen, Janne, Virtanen, Juha, . . . Naatenen, Risto. (1999). Analysis of speech sounds is left-hemisphere predominant at 100± 150

ms after sound onset. *NeuroReport, 10*(5), 1113-1117.

Ritter, W., Sussman, E., Deacon, D., Cowan, N., & Vaughan, H.G.Jr. (1999). Two cognitive systems simultaneously prepared for opposite events. *Psychophysiology, 36, 835-839., 36*, 835-839.

Rogalsky, C., Love, T., Driscoll, D., Anderson, S.W., & Hickok, G. (2011). Are mirror neurons the basis of speech perception? Evidence from five cases with damage to the purported human mirror system. *Neurocase, 17*(2), 178-187.

Rogalsky, C., Pitz, E., Hillis, A.E., & Hickok, G. (2008). Auditory word comprehension impairment in acute stroke: Relative contribution of phonemic versus semantic factors. *Brain and Language, 107*(2), 167-169.

Schafer, V.L., Schwartz, R.G., & Martin, B. (2011). Evidence of deficient central speech processing in children with specific language impairment: The T-complex. *Clinical Neurophysiology, 122*, 1137-1155.

Scherg, M., & von Cramon, D.Y. (1986). Evoked dipole source potentials of the human auditory cortex. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section, 65*(5), 344-360.

Schofield, T.M., Iverson, P., Kiebel, S.J., Stephan, K.E., Kilner, J.M., Friston, K.J., . . . Leff, A.P. (2009). Changing meaning causes

coupling changes within higher levels of the cortical hierarchy. *Proceedings of the National Academy of Sciences, 106*(28), 11765-11770.

Schofield, T.M., Penny, W.D., Stephan, K.E., Crinion, J.T., Thompson, A.J., Price, C.J., & Leff, A.P. (2012). Changes in Auditory Feedback Connections Determine the Severity of Speech Processing Deficits after Stroke. *The Journal of Neuroscience, 32*(12), 4260-4270.

Sebastien-Galles, N., & Baus, C. (2005). On the Relationship Between Perception and Production in L2 Categories. In A. Cutler (Ed.), *Twenty-first century psycholinguistics: Four cornerstones* (pp. 279-292). New York: Erlbaum.

Segalowitz, N., & Hulstijn, J. (2005). Automaticity in Bilingualism and Second Language Learning. In J. Kroll & A. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 371-388). Oxford, UK: Oxford University Press.

Segalowitz, N., & Segalowitz, S. (1993). Skilled performance, practice, and the differentiation of speed-up from automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics, 14*, 369-385.

Segalowitz, N., Segalowitz, S., & Wood, A. (1998). Assessing the development of automaticity in second language word recognition. *Applied Psycholinguistics, 19*, 53-67.

Sharma, A., & Dorman, M. (2000). Neurophysiologic correlates of cross-language phonetic perception. *Journal of the Acoustical Society of America, 107*, 2697-2703.

Skoe, Erika, & Kraus, Nina. (2010). Auditory brainstem response to complex sounds: a tutorial. *Ear and Hearing, 31*(3), 302-324.

Slabu, Lavinia, Escera, Carles, Grimm, Sabine, & Costa-Faidella, Jordi. (2010). Early change detection in humans as revealed by auditory brainstem and middle-latency evoked potentials. *European Journal of Neuroscience, 32*(5), 859-865.

Stapells, D.R. (2002). Cortical event-related potentials to auditory stimuli. . In J. Katz, R. F. Burkard & L. Medwetsky (Eds.), *Handbook of clinical audiology* (pp. 378-406). Philadelphia, PA: Lippincott, Williams and Wilkins.

Steinschneider, M., Liegeois-Chauvel, C., & Brugge, J. F. . (2011). Auditory evoked potentials and their utility in the assessment of complex sound processing. In C. Schreiner & J. Winer. (Eds.), *The Auditory Cortex.* (pp. 535-559): Springer.

Stephan, K.E., Penny, W.D., Daunizeau, J., Moran, R.J., & Friston, K.J. (2009). Bayesian model selection for group studies. *NeuroImage, 46*(4), 1004-1017.

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception and Psychophysics, 36*, 131-145.

Sussman, E. (2007). A new view on the MMN and attention debate: Auditory context effects. . *Journal of Psychophysiology, 21*(3-4), 164-175.

Sussman, E., Sheridan, K., Kreuzer, J., & Winkler, I. . (2003). Representation of the standard: Stimulus context effects on the process generating the mismatch negativity component of event-related brain potentials. *Psychophysiology, 40*(465-471).

Sussman, E., & Winkler, I. (2001). Dynamic sensory updating in the auditory system. *Cognitive Brain Research, 12*(3), 431-439.

Swaminathan, J., Krishnan, A., Gandour, J.T., & Xu, Y. (2008). Applications of static and dynamic iterated rippled noise to evaluate pitch encoding in the human auditory brainstem. *IEEE Transactions on Biomedical Engineering, 55*(1), 281-287.

Teki, S., Barnes, G.R., Penny, W.D., Iverson, P., Woodhead, Z.V.J., Griffiths, T.D., & Leff, A.P. (2013). The right hemisphere supports but does not replace left hemisphere auditory function in patients with persisting aphasia. *Brain, 136*(6), 1901-1912.

Tervaniemi, M, & Hugdahl, K. (2003). Lateralization of auditory-cortexfunctions. *Brain Research Reviews, 43*, 231-246.

Tonnquist-Uhlen, I., Ponton, C.W., Eggermont, J.J., Kwong, B., & Don, M. (2003). Maturation of human central auditory system

activity: the T-complex. *Clinical Neurophysiology, 114*(4), 685-701.

Ulanovsky, N., Las, L., & Nelken, I. (2003). Processing of low-probability sounds by cortical neurons. *Nat Neurosci, 6*(4), 391-398.

Uther, M., Giannakopoulou, A., & Iverson, P. (2012). Hyperarticulation of vowels enhances phonetic change responses in both native and non-native speakers of English: Evidence from an auditory event-related potential study. *Brain Research, 1470*(0), 52-58.

Vaughan Jr, H.G., & Ritter, W. (1970). The sources of auditory evoked responses recorded from the human scalp. *Electroencephalography and Clinical Neurophysiology, 28*(4), 360-367.

Wagner, M., Shafer, V.L., Martin, B., & Steinschneider, M. (2013). The effect of native-language experience on the sensory-obligatory components, the P1–N1–P2 and the T-complex. *Brain Research, 1522*(0), 31-37.

Watkins, K. E., Strafella, A. P., & Paus, T. . (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia, 41*(8), 989-994.

Wilson, S. M., Saygin, A. E. P., Sereno, M. I., & Iacoboni, M. . (2004). Listening to speech activates motor areas involved in

speech production. *Nature Neuroscience, 7*(7), 701-702.

Winkler, I., Cowan, N., Csepe, V., Czigler, I., & Näätänen, R. (1996). Interactions between Transient and Long-Term Auditory Memory as Reflected by the Mismatch Negativity. *Journal of Cognitive Neuroscience, 8*(5), 403-415.

Winkler, I., & Czigler, I. (1998). Mismatch negativity: deviance detection or the maintenance of the 'standard'. *NeuroReport, 9*(17), 3809-3813.

Winkler, I., Karmos, G., & Näätänen, R. (1996a). Adaptive modeling of the unattended acoustic environment reflected in the mismatch negativity event-related potential. *Brain Research, 742*(1–2), 239-252. doi: http://dx.doi.org/10.1016/S0006-8993(96)01008-6

Winkler, I., Karmos, G., & Näätänen, R. (1996b). Adaptive modeling of the unattended acoustic environment reflected in the mismatch negativity event-related potential. *Brain research, 742*, 239-252.

Winkler, I., Kujala, T., Tiitinen, Hannu, Sivonen, Paivi, Alku, Paavo, Lehtokoski, Anne, . . . Naatanen, Risto. (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology, 36*, 638-642.

Winkler, I., Lehtokoski, A., Alku, P., Vainio, M., Czigler, I., Csépe, V., . . . Näätänen, Risto. (1999). Pre-attentive detection of

vowel contrasts utilizes both phonetic and auditory memory representations. *Cognitive Brain Research, 7*, 357-369.

Zatorre, R.J., Belin, P., & Penhune, V.B. (2002). Structure and function of auditory cortex: music and speech. *Trends in Cognitive Sciences, 6*(1), 37-46.

Zevin, J., Datta, H., Maurer, U., Rosania, K. A., & McCandliss, B. D. . (2010). Native language experience influences the topography of the mismatch negativity to speech. *Frontiers in Human Neuroscience, 4*.